22411/B

EXPERIMENTAL DESIGN IN PSYCHOLOGICAL RESEARCH

By
Edwards, A.L.

BOOKS BY ALLEN L. EDWARDS

Experimental Design in Psychological Research, *Revised Edition*

Statistical Methods for the Behavioral Sciences

Statistical Analysis, *Revised Edition*

Workbook to Accompany the Revised Edition of Statistical Analysis

Social Desirability Variable in Personality Assessment and Research

Techniques of Attitude Scale Construction

TO THOSE STUDENTS

who will some day make their contribution
to psychology and the behavioral sciences
by research and experimentation

PREFACE

In preparing this revision of *Experimental Design in Psychological Research*, I have been guided by the same principles I followed in writing the first edition. I have tried to write a book which can be understood by those familiar with elementary statistical analysis and who have a working knowledge of algebra.

A number of hypothetical sets of data are interspersed with the results of actual experiments. I have attempted to arrange the problems at the end of each chapter so that some illustrate by fairly easy computations the analyses described in the text. Some problems requiring more prolonged calculations have also been included. The problems are in all cases modeled after the methods presented in the chapters. I have also included in the problems at various times a brief discussion of a particular point which the problem has been designed to illustrate. Answers to the problems are given in the appendix.

I am indebted to Sir Ronald A. Fisher and to Messrs. Oliver Boyd Ltd., Edinburgh, for permission to reprint Tables IV, V, and VI from their book *Statistical Methods for Research Workers*. Table I is reproduced by permission of Messrs. Kendall and Smith and the Royal Statistical Society. Table VIII has been reproduced from Professor Snedecor's book, *Statistical Methods*, by permission of the author and his publisher, the Iowa State College Press. Additional values of t were also taken from Professor Snedecor's book by permission. Portions of Table II have been taken from *Handbook of Statistical Nomographs, Tables, and Formulas* by permission of Drs. J. W. Dunlap and A. K. Kurtz and their publishers, the World Book Company.

I am also indebted to Maxine Merrington and Catherine M. Thompson and *Biometrika* for permission to reproduce the entries in Table IX. H. Leon Harter kindly made available to me, in advance of publication, Tables Xa–Xe. Tables XIIa–XIId were developed by Charles W. Dunnett and are reproduced by permission of the *Journal of the American Statistical Association*.

To the American Psychological Association, the American Statistical Association, the Biometrics Society, the Royal Statistical Society, the

British Psychological Society, the University College of London, Williams and Wilkins Company, and Warwick and York, Inc., I am indebted for permission to quote material and to make use of data published in various professional journals.

Dr. Doris M. Lee, of the University of London Institute of Education, was of assistance in checking my computations Dr. Leonard S. Kogan, Director of Institute of Welfare Research, Community Service Society, read the original draft of the book, and Dr. David A. Grant, Professor of Psychology at the University of Wisconsin, read Chapters 14 and 15. Their suggestions and comments have been helpful in many ways.

Finally, I should like to express my appreciation to the University of Washington for granting me sabbatical leave while I was writing the book, and to the John Simon Guggenheim Memorial Foundation for awarding me a Fellowship at the time the writing was completed. The Guggenheim Fellowship made possible a year of research and study in London, during which time I was able to read and correct proofs without also having to be concerned with teaching responsibilities.

January, 1960 ALLEN L. EDWARDS

CONTENTS

Preface vii

1. THE NATURE OF RESEARCH 1

Introduction, 1; Observations and Variables, 3; Stimulus Variables, 5; Behavioral Variables, 6; Organismic Variables, 7; Research in Psychology, 9; Experiments, 9; *Questions and Problems,* 11

2. PRINCIPLES OF EXPERIMENTAL DESIGN 13

Introduction, 13; The Farmer from Whidbey Island, 14; The First Experiment, 15; Significance Levels, 18; Two Types of Errors, 19; Experimental Controls, 19; The Importance of Randomization, 21; A Limitation in the First Experiment, 22; The Second Experiment, 23; *Questions and Problems,* 25

3. BINOMIAL POPULATIONS IN RESEARCH 28

Introduction, 28; Random Selection and Random Samples, 28; Probabilities of Possible Outcomes, 30; The Binomial Population, 31; Statistics and Parameters, 33; Frequency and Sampling Distributions, 34; Example of a Finite Population Model, 35; The General Case for Samples from Finite Populations, 36; Example of an Infinite Population Model, 38; The General Case for Samples from Infinite Populations, 39; Finite Population Correction Factor, 40; Binomial Expansion, 40; Applications of the Models in Research, 41; *Questions and Problems,* 42

4. APPROXIMATION OF THE PROBABILITIES ASSOCIATED WITH SAMPLING FROM A BINOMIAL POPULATION 43

Introduction, 43; The Unit Normal Distribution, 43; Two Examples of Approximating Binomial Probabilities, 46; Testing Null Hypotheses, 48; Test of Significance of p, 49; The Matching Problem, 49; Significance of the Difference Between Two Proportions, 51; Exact Test for the Difference Between Two Proportions, 55; Test for the Difference Between

Two Proportions When the Same Subjects Are Tested Twice, 57; *Questions and Problems*, 60

5. TESTS OF SIGNIFICANCE WITH THE χ^2 DISTRIBUTION 63

Introduction, 63; One Sample with c Classes, 64; Two or More Samples with c Classes, 65; Two or More Samples with $c = 2$ Classes, 67; Two Samples with $c = 2$ Classes, 69; Test of Technique, 71; χ^2 with More than 30 d.f., 73, *Questions and Problems*, 74

6. SIGNIFICANCE TESTS FOR THE CORRELATION COEFFICIENT 77

Introduction, 77; Sampling Distribution of r, 77; The t Test of the Hypothesis of Zero Correlation, 78; Table of Significant Values of r, 79; The z' Transformation for r, 79, Significance of the Difference Between Two r's, 82; Test of Homogeneity of k Values of r, 83; Significance of the Difference Between Nonindependent r's, 85; *Questions and Problems*, 85

7. THE t TEST FOR MEANS 86

Introduction, 86; Sampling Distribution of the Mean, 87; The t Distribution, 88; Confidence Limits for the Mean, 89; Difference Between Two Means, 90; Standard Error of the Difference Between Two Means, 92; Confidence Limits for a Mean Difference, 93; Test of Significance of a Mean Difference, 94; The Null Hypothesis and Alternatives, 94; Number of Observations, 97; Randomization, 100; *Questions and Problems*, 101

8. HETEROGENEITY OF VARIANCE AND THE t TEST 104

Introduction, 104; The F Distribution, 104; Testing for Homogeneity of Variance, 105; Heterogeneity of Variance with $n_1 \neq n_2$, 106; Heterogeneity of Variance with $n_1 = n_2$, 107; Conditions Making for Heterogeneity of Variance, 108; Assumptions of the t Test, 111; *Questions and Problems*, 114

9. INTRODUCTION TO THE ANALYSIS OF VARIANCE 117

Introduction, 117; Calculations for a Randomized Groups Design, 118; Nature of the Sum of Squares in a Randomized Groups Design, 121; Mean Squares and the Test of Significance in a Randomized Groups Design, 123; Heterogeneity of Variance, 125; Transformations of Scale, 128; Further Comments on the F Test and Heterogeneity of Variance, 131; *Questions and Problems*, 132

10. MULTIPLE COMPARISONS IN THE ANALYSIS OF VARIANCE 136

Introduction, 136; Duncan's New Multiple Range Test, 136; Orthogonal Comparisons of Treatment Means, 140; Orthogonal Comparisons of Treatment Sums, 144; Trend Analysis, 148, Dunnett's Test for Comparisons with a Control, 152, Scheffé's Test for Multiple Comparisons, 154; *Questions and Problems*, 157

11. THE RANDOMIZED BLOCKS DESIGN 158

Introduction, 158; Example of a Randomized Blocks Design, 159; Sums of Squares in the Randomized Blocks Design, 162; Variables Used in Forming Blocks, 164; Nonadditivity, 165; Randomized Blocks with $k = 2$ Treatments, 169; *Questions and Problems*, 172

12. THE $2 \times 2 \times 2$ FACTORIAL EXPERIMENT 175

Introduction, 175; A $2 \times 2 \times 2$ Factorial Experiment, 175; Two-Part Analysis of Variance, 177; Partitioning the Treatment Sum of Squares, 179; Meaning of the Main Effects, 183; The Interaction Effects, 184; Summary of the Conclusions, 188; Orthogonal Comparisons, 189; Notation and Sums of Squares, 191; Further Discussion of Interactions, 192; Other 2^n Factorial Experiments, 196; Advantages of Factorial Experiments, 197; *Questions and Problems*, 199

13. FACTORIAL EXPERIMENTS: FURTHER CONSIDERATIONS 201

Introduction, 201; A $4 \times 3 \times 2$ Factorial Experiment, 201; Direct Calculation of a Three-Factor Interaction, 207; Many Factors with Many Levels, 211; Factorial Experiments with Randomized Blocks, 213; Organismic Variables as Factors, 215; An Experiment with an Organismic Factor, 217; *Questions and Problems*, 219

14. TREND ANALYSIS 224

Introduction, 224; Trial Means: One Standard Condition, 226; Trial Means: Different Treatments, 227; Trial Means: A Treatment Factor and an Organismic Factor, 233; Trend Analysis of the Over-all Stage Sums, 239; Linear Components of Interactions with Stages, 240; Quadratic Components of Interactions with Stages, 243; Tests of Significance of Linear and Quadratic Components, 245; Trend Analysis of the Drug Experiment, 247; *Questions and Problems*, 250

15. LATIN SQUARE DESIGNS 254

Introduction, 254; Analysis of Variance of the Latin Square Design, 255; General Equation for the Latin Square, 257; Randomization Procedures, 258; Replication with Independent Squares, 259; Replication of the Same Square, 265; Latin Squares and Fractional Replication, 270; The 2×2 Latin Square, 271; Carry-Over Effects, 274; Graeco-Latin Squares, 276; *Questions and Problems*, 277

16. THE ANALYSIS OF COVARIANCE FOR A RANDOMIZED GROUPS DESIGN 281

Introduction, 281; Product Sums, 281; Relationship Between X and Y in the Absence of Treatment Effects, 283; Relationship Between X and Y When Treatment Effects are Present, 285; Sums of Squares and Product Sums for a Randomized Groups Design, 286; Variation Within Each Group About the Regression Line for the Group, 288; Variation Within Groups About a Common Regression Line with Slope b_w, 290; Test of Significance of Differences Between the Group Regression Coefficients, 291; Test of Significance for the Treatment Means, 292; Nonlinear Relationship Between X and Y, 294; Analysis of Difference Measures, 295; A Randomized Blocks Design as an Alternative to the Analysis of Covariance, 296; Several Supplementary Measures, 298; Analysis of Covariance and Other Experimental Designs, 299; *Questions and Problems*, 299

17. ANALYSIS OF VARIANCE MODELS AND EXPECTATIONS OF MEAN SQUARES 301

Introduction, 301; Model II: Expectations of Mean Squares, 302; Expectations of Mean Squares for a Mixed Model, 303; Expectations of Mean Squares in a Randomized Blocks Design, 306; Expectations of Mean Squares in Split-Plot Designs, 309; Expectations of Mean Squares in the Latin Square Design, 311; *Questions and Problems*, 313

REFERENCES 315

LIST OF FORMULAS 321

APPENDIX 331

Table I. Table of Random Numbers 332

Table II. Table of Squares, Square Roots, and Reciprocals of Numbers from 1 to 1,000 337

Table III. Areas and Ordinates of the Normal Curve in Terms of x/σ 350
Table IV. Table of χ^2 360
Table V. Table of t 361
Table VI. Values of the Correlation Coefficient for Different Levels of Significance 362
Table VII. Table of z' Values for r 363
Table VIII. The 5 (Roman Type) and 1 (Boldface Type) Per Cent Points for the Distribution of F 364
Table IX. The 25, 10, 2.5, and 0.5 Per Cent Points for the Distribution of F 368
Table Xa. Significant Studentized Ranges for Duncan's New Multiple Range Test with $\alpha = .10$ 372
Table Xb. Significant Studentized Ranges for Duncan's New Multiple Range Test with $\alpha = .05$ 373
Table Xc. Significant Studentized Ranges for Duncan's New Multiple Range Test with $\alpha = .01$ 374
Table Xd. Significant Studentized Ranges for Duncan's New Multiple Range Test with $\alpha = .005$ 375
Table Xe. Significant Studentized Ranges for Duncan's New Multiple Range Test with $\alpha = .001$ 376
Table XI. Coefficients for Obtaining the Linear and Quadratic Components of the Treatment Sum of Squares When Treatments Are Equally Spaced 377
Table XIIa. Table of t for One-Sided Comparisons Between k Treatment Means and a Control for a Joint Confidence Coefficient of $P = 95$ Per Cent 378
Table XIIb. Table of t for One-Sided Comparisons Between k Treatment Means and a Control for a Joint Confidence Coefficient of $P = 99$ Per Cent 379
Table XIIc. Table of t for Two-Sided Comparisons Between k Treatment Means and a Control for a Joint Confidence Coefficient of $P = 95$ Per Cent 380
Table XIId. Table of t for Two-Sided Comparisons Between k Treatment Means and a Control for a Joint Confidence Coefficient of $P = 99$ Per Cent 381
Table XIII. Table of Four-Place Logarithms 382

ANSWERS TO PROBLEMS 385

AUTHOR INDEX 391

SUBJECT INDEX 393

EXPERIMENTAL DESIGN
IN PSYCHOLOGICAL RESEARCH

1

THE NATURE OF RESEARCH

INTRODUCTION

Incidental observation is frequently important in initiating research, for such observation may motivate us to formulate hypotheses, to state problems, to ask questions. We may then seek answers to these questions by planned or systematic observation which characterizes research and experimentation. In research, we do not haphazardly make observations of any and all kinds, but rather our attention is directed toward those observations which we believe to be relevant to questions we have previously formulated. The objective of research, as recognized by all sciences, is to use observation as a basis for answering questions of interest.[1]

Even when observations are made systematically, the raw unclassified observations are often of such a nature that in their original form they do not lend themselves to an obvious interpretation with respect to the questions we have posed. We, therefore, resort to techniques which will reduce the observations to a more manageable form. These techniques involve classifying and operating upon the observations to reduce them to frequencies, proportions, means, variances, correlation coefficients, and other statistical measures. On the basis of these statistical measures we hope to be able to draw certain conclusions or inferences which will bear upon questions of interest.

To consider a simple example, suppose that for some reason or another we have become interested in the question of whether or not a particular coin is unbiased. By unbiased we mean that if the coin is tossed in a random fashion it is equally probable that it will fall heads or tails. In an attempt to obtain an answer to the question raised, we may undertake systematic observation of the outcomes of a specified number of tosses of the coin. A single observation will consist of the result of a single toss; that is, we observe and record whether the coin falls heads or tails.

[1] The brief discussion of research and experimentation in the behavioral sciences given in this chapter obviously does not do justice to the subject. For a much more complete treatment of research methods and techniques, see Lindzey (1954), Festinger and Katz (1953), Jahoda, Cook, and Deutsch (1959), Brown and Ghiselli (1955), and Underwood (1957).

If we make 100 observations, they may be recorded in the form HTTHTT . . . H, with each H and T being the record of a single observation But this series of observations, in unreduced form, is not easy to interpret with respect to the question of interest. We can reduce the observations, however, by counting the number of H's and the number of T's and these two frequencies will summarize succinctly the complete set of observations.

It may be intuitively obvious to us that, if the coin is unbiased, the two frequencies should be approximately equal and that any departure from equality would provide *some* evidence against the notion that the coin is unbiased. An important problem in research is how to evaluate objectively the evidence provided by any given set of observations. We shall see later that there are techniques for doing this and that these techniques require additional operations upon the two observed frequencies.

The example cited characterizes fairly well the nature of much research. One or more questions are formulated. Systematic observation is then made of things believed to be relevant to the questions. Having made a systematic series of observations, the observer then reduces these to a limited number of statistical measures which provide a summary description of the complete set. By means of further operations upon the descriptive measures, the evaluation of the evidence provided by the observations, with respect to the questions of interest, is placed on an objective basis.

A number of additional aspects of research are also illustrated by the example. How many observations should be made? Obviously, only one observation would not provide us with adequate information in the present instance. If an insufficient number of observations are made, we shall not be much better off than with none. On the other hand, if we make more observations than needed, we shall have wasted time and energy that might be more fruitfully spent in other endeavors. We shall later discuss techniques of assistance in estimating the number of observations adequate for stated objectives.

We have mentioned also the relevance of the observations to the questions of interest. In the case at hand, the relevance of the observations made may seem obvious. We do not, for example, direct our attention to observing the position of the coin on the floor after each toss, but rather to observing whether the coin falls heads or tails. We observe the latter because we believe these observations are pertinent to the question we have asked. In other research problems, the relevance of the observations made to the question of interest may not always be so clear-cut. Whether the observations to be made in a given research problem are relevant to the question they are supposed to answer must always be given serious consideration in the planning of the research.

The manner in which the coin is tossed is also of importance. If the coin is tossed in a systematic fashion, this may result in our observing an excess of heads over tails or vice versa. The observations then made would bear not upon the nature of the coin—the problem of interest—but rather upon the nature of the toss. We have indicated that the method of tossing should be such as to introduce randomness in the observations so that, if the coin is unbiased, heads or tails will be equally probable.

Furthermore, although we have indicated that techniques are available for making the evaluation of the evidence provided by a set of observations objective, this does not mean that we shall always be correct in the inference made simply because it is made on an objective basis. We may, for example, conclude that the coin is biased when, in fact, it is not. Or we may fail to conclude that the coin is biased, when, in fact, it is. These two kinds of errors also require consideration in the planning of research.

In this book, we shall be concerned with the various problems, some of which have been briefly described, involved in planning research and in designing experiments. We shall also be concerned with the analysis of observations made in the course of research. If a research study is well planned or an experiment is well designed, the methods of analysis to be applied to the observations will have been given consideration in the planning stage. Only in this way can one have some assurance that the observations to be made will enable one to obtain answers to the questions of interest. It would obviously be frustrating to any research worker if he were to make a large number of observations, only to find that they could not be analyzed in any way which would answer the questions he was interested in. The only safe way in which to avoid this kind of frustration is to plan the analysis *before* the observations have been made.

OBSERVATIONS AND VARIABLES

We have stressed the importance of observations in research. The things which are observed are called *variates* or *variables*. In the example cited earlier, the variable, the thing observed, was the face of the coin. Any particular observation is called a *value* of the variable. For the face of a coin there are only two possible values of the variable, heads and tails. The value of a variable indicates the class to which an observation is to be assigned. If a coin falls heads, we consider that observation and all others with the same value as belonging to the same class. In order for something we are observing to be considered a variable, we must have at least two possible classes of observations. The classes must also be mutually exclusive; that is, any given observation can be assigned to only one of the available classes. By a variable, therefore, we shall mean anything that we

can observe and of such a nature that each single observation can be classified into one of a number of mutually exclusive classes.

Any given variable can also be described as quantitative or qualitative, depending upon the nature of the observations made. We shall examine first some of the characteristics of quantitative variables and then those of qualitative variables.

Quantitative Variables

A given series of observations may be obtained by measurement, as, for example, when a stimulus under observation consists of a beam of light and the intensity of the light is varied and measured. The differences in the intensity of the light are matters of degree and these may be measured, giving rise to a series of *quantitative* observations. This series is also *continuous* in that we could theoretically increase or decrease the intensity of the light by infinitesimal amounts. It is also true that no matter how fine we might make the differences in intensity, they could always theoretically be made finer. That they are not may simply be a function of our measuring instrument, or it may be due to the fact that we had no need for more precise observations. Consequently, we may note that observations of a quantitative and continuous variable are *approximate* and not exact.

If we were interested in a problem involving a light stimulus, we might keep the intensity of the light constant and vary the number of times the light was presented. A series of such observations is quantitative but not continuous. The number of times that the light can be presented can increase or decrease only by integral numbers and not by indefinitely small fractions. We may present the light 4 times or 5 times, but not by any decimal fraction falling between 4 and 5. This series of observations would be quantitative and *discrete*. Quantitative data obtained by counting are often referred to as *enumeration* data or *frequency* data. Such data differ from the continuous data obtained by measurement in that they are *exact*. No approximation is involved, for example, in counting the number of times the stimulus was presented; if no error occurred in making the observations, the frequency of presentation is known exactly.

The difference between a continuous and a discrete series of observations may be illustrated also in terms of a response variable. We might observe the number of times a particular response occurs to the light stimulus. For example, in a conditioning experiment, if the light stimulus has always been followed by an electric shock to the finger in a preliminary series of trials, we might count the number of times the finger is flexed to the light stimulus when it is presented alone during a series of critical trials. This result would give us a quantitative and discrete series of observations of the response. On the other hand, we might record our observations of

the finger response by measuring the amplitude of the finger flexion and thus obtain a quantitative and continuous series of observations.

Qualitative Variables

Our interest might be in the ease with which we may establish conditioned finger flexion in the right hand as compared with the conditioning of an eye blink. The response variable which we observe would thus be of a *qualitative* kind, that is, difference in response of the hand and response of the eye is a matter of kind, not degree. These two qualitatively different responses, of course, might each be recorded as a quantitative continuous or a quantitative discrete series of observations.

Qualitative variables are also often described as *unordered* variables. If observations of a thing differ with respect to amount or degree, then the observations can be arranged and classified in such a way that one class can be said to represent more, or a greater degree, of a variable than another class. For example, suppose that a variable under observation is height. Then an observation that has the value 74 inches can be said to represent a greater degree of height than one which has the value 70 inches. With qualitative variables, the various classes of observations have no ordered relations, since the differences between the classes are matters of kind rather than degree.

STIMULUS VARIABLES

Although psychology has, on occasion, been defined as the science of behavior, it is obvious that behavior does not occur in a vacuum but always in a particular setting or environment. The general class of things we observe that relate to the environment, situation, or conditions of stimulation, we shall refer to as *stimulus* variables. The stimulus variables in a psychological experiment may consist of relatively simple things, such as electric shock, light, sound, pressure, or temperature. These may be quantified by measuring the physical intensity of the stimulus.

There are other stimulus variables in which the psychologist is interested for which we have no measures corresponding to physical intensity. These may consist of problem-solving situations, motor conflict situations, social situations, and so forth, and these are relatively more difficult to quantify. Indeed, in much research we can only say that the variations in stimulation which we are interested in consist of complex combinations of stimuli differing in kind rather than in degree. We shall refer to any such differences in the conditions of stimulation, in a given experiment, as differences in *treatments*. The term treatment will thus be used to refer to a particular set of stimulus or experimental conditions.

In a given experiment, for example, we may be interested in certain behavior of subjects when they are involved with an "authoritarian" leader and in the behavior of other subjects when they are involved with a "democratic" leader. Other than the behavior of the leader, we may attempt to keep all other aspects of the stimulus situation constant. Assuming we are successful in this respect, we must still recognize that the differential role-playing activities of the leader may be expected to result in complex differences in the two situations. The difference between the two situations, in other words, may not be in one dimension or variable but rather in a set of many differences in many variables. The term treatments is a useful one since it can be used in connection with differences between complex sets of experimental or stimulus conditions, and also in reference to differences in experimental conditions where a single variable is involved.

BEHAVIORAL VARIABLES

By a *behavioral* variable we shall mean any action of an organism. At one extreme, these actions may consist of relatively simple responses such as finger flexions, eye blinks, knee jerks, and so forth. At the other extreme, we have such complex behavior patterns as those involved in typewriting, tennis playing, problem solving, or perhaps even the more complicated behavior involved in aggression, dominance, leadership, and social adaptability.

We have already noted that a simple stimulus variable is quantified more readily than a relatively complex one. This is true for behavioral variables also. Thus we have apparatus for recording and quantifying in a continuous series a relatively simple response such as the dilation of the pupil of the eye or the flexion of a finger. As responses become more complex, we may still have quantitative observations but they tend to be discrete rather than continuous. In measuring typewriting skill, for example, we may record the number of words typed per unit of time. In studying stylus maze learning of the human subject, we may count the number of errors made per trial. We may count the number of aggressive responses made by a child in a social situation or the number of times the child withdraws in response to the advances of another child. In each instance we have a quantitative series of observations, but they are discrete and not continuous.

We attempt, however, to obtain a continuous series of observations of the more complicated response patterns by a variety of techniques. We devise rating scales and ask judges to rate the degree of aggressiveness or submissiveness exhibited by a child in a given situation. Although these ratings may be made on a discrete scale, we often assume that they repre-

sent a continuous series. Thus, if the rating scale contains but five discrete categories, ranging from 1 to 5, with 1 representing a minimum degree of aggressiveness and 5 representing a maximum, and a given child obtains a rating of 3 on the scale, we may take this rating to represent an interval. For example, we do not treat the rating as an exact figure of 3, but instead assume that the rating represents an interval ranging perhaps from 2.5 up to 3 5. In other words, the rating of 3 is taken as an approximate measurement rather than as an exact, discrete value. We do this because we grant the logic of recognizing that aggressiveness does not increase or decrease by successive or discrete values, but rather that human beings may be thought of as occurring on a continuum separated by degrees of aggressiveness. That we are not able to locate subjects more precisely on this continuum, but only in terms of an interval ranging .5 of a unit below and .5 of a unit above the recorded values, is a result of our technique of observation. The apparent discreteness of our observations is not believed to be an inherent characteristic of the variable observed.

In the same way, we often treat scores on psychological tests as representing approximate measurements rather than exact discrete values. We may count the number of items responded to in a particular way on a test and, although these counts are discrete, we again recognize that the discreteness is an artifact of our method of observation. We might just as well have assigned fractional values to the various items in the test and thus have obtained scores which were separated by smaller values than those we have obtained by assigning the value of unity to each item responded to in a particular manner. The method of assigning weights to items in a test is often an arbitrary matter, and hence, though it may be more convenient to assign simple weights of unity, we treat the sum of these weights, the score, as belonging to a continuous series. A score of 18, for example, is treated as if it represented an interval ranging from 17.5 up to 18.5 and thus is considered as an approximate value in a continuous series rather than an exact value of 18.

ORGANISMIC VARIABLES

Organismic variables arise from ways in which organisms may be classified and from the observations and measurements of physical, physiological, and psychological characteristics of organisms. For example, we may measure the heights or weights of a group of individuals, and the resulting measurements would constitute a quantitative and continuous series of observations. These observations do not correspond to response variables or stimulus variables, but they may be conveniently described as *organismic* variables. They are characteristic ways in which the particular group of organisms under observation vary. Similarly, organisms may be

classified as to the color of their hair or eyes, and these classifications would constitute a series of qualitative observations.

We may also classify individuals in terms of their educational levels. Some will have no schooling, some will be grammar-school graduates, some high-school graduates, and some college graduates. This series of observations might be arbitrarily quantified by assigning 1 to the lowest level of education, 2 to the next lowest, and so on. Or we might simply record the number of years of schooling for each individual.

Individuals may also be classified according to their sex, and this gives us a qualitative series. But this qualitative series is often, for reasons of convenience, identified by assigning a value of 0 to one sex and a value of 1 to the other.

Rats in an experiment on learning may be characterized as hungry or thirsty, and we have here a qualitative series of observations or characteristics of the rats, for this is a difference in kind not in degree. But we may quantify within these qualitative differences by designating some rats as more hungry than others and some rats as more thirsty than others. Such designations would probably be based, of course, upon prior knowledge that some of the rats had not been fed or given water in 24 hours, whereas others had been fed or permitted to drink water as recently as 12 hours before the experiment. With this knowledge we might arbitrarily assign weights of 0 and 1 to the 12- and 24-hour periods of deprivation, respectively. Or we might use the time intervals 12 and 24 hours as our quantitative measures.

In studying behavior characteristics of young children we may classify them into two groups: those who were breast-fed and those who were bottle-fed. In studying the responses of adults to some social issue, we may find it convenient to classify them into those who voted the Democratic ticket and those who voted the Republican ticket. In still other cases, we may classify individuals as to the size of the town in which they were raised, or in terms of the size of the high school from which they graduated, or on the basis of nationality backgrounds. Each of these various classifications, in a given research problem, may be regarded as an organismic variable.

In research, frequent use is also made of *response-inferred* organismic variables. By a response-inferred organismic variable is meant a classification based upon prior observation of response. A person's IQ, for example, is determined by observing his response to a standardized testing situation. It is convenient, however, in many cases, to regard IQ as something that is associated with the organism, that is, as an organismic variable. As another example of a response-inferred organismic variable, it is not uncommon to refer to one group of subjects in a given research problem as the "anxious" group and another group as the "nonanxious" group. This classification is often based upon prior observations of response to some test of anxiety.

RESEARCH IN PSYCHOLOGY

It is the concern of psychologists and other scientists who are interested in the behavior of organisms to describe and study stimulus, response, and organismic variables. Much of psychological research is concerned with attempts to improve our methods of description of these variables by devising apparatus and developing techniques for more precise measurement of them. In the hands of other psychologists these devices are used to make systematic observations of variables of interest.

As we have stated earlier, systematic observation is undertaken in an attempt to obtain answers to questions in which we are interested. In some cases the questions of interest have to do with the accurate description of a group of subjects with respect to one or more variables. An instructor, for example, may be interested in how the intelligence test scores for one of his classes are distributed. His systematic observations might consist of the score of each of his students on some standardized test of intelligence. These observations may be reduced—classified—so that he has available for each score the frequency with which it was observed. He may also be interested in finding out what the average score is for his class and something about the range or spread of scores.

In other cases, the questions of interest may concern the degree of association or relationship between two variables. For example, the same instructor may also be interested in determining whether or not the intelligence test scores of the students are in any way related to or associated with scores on a standardized achievement test. Do, for example, students with high intelligence test scores also tend to obtain high scores on the achievement test, while those with low scores on the intelligence test also tend to obtain low scores on the achievement test?

EXPERIMENTS

In certain instances it is possible for an investigator to vary quantitatively one variable, usually a stimulus variable, and to study the behavior of groups of subjects under each value of the variable. As an example, we might vary the size of type in which a list of words is printed. Subjects are assigned to a given type size and the words for each type size are exposed at a constant rate. The observations obtained for each variation in the stimulus conditions might be the number of words correctly recognized by each subject. If the average number of words correctly recognized for each type size is obtained, we may then determine the relationship between these averages and the type size. We may, for example, be interested in finding out whether the average number of words recognized increases as

the type size is increased. Furthermore, we may be interested in determining whether the relationship between these two variables is linear or not.

In the case of the instructor interested in the relationship between scores on the intelligence test and scores on a standardized achievement test, the instructor is not able to control or manipulate the variables in which he is interested. For example, he has no control over the intelligence test scores nor over the achievement test scores of his subjects. The values of these observations are fixed or determined by each subject and can not be manipulated or changed directly by the instructor. In the case just described, however, one of the variables was directly under the control of the investigator. This variable was the stimulus variable, that is, the type size. The investigator can vary or alter this variable in the manner described. When certain variables can be controlled or manipulated directly in a research problem by the investigator, the research procedure is often described as an *experiment.*

The variables over which the investigator has control are called the *independent* variables. They are those which the investigator himself manipulates or varies. As the independent variables are changed or varied, the investigator observes other variables to see whether they are associated with or related to the changes introduced. These variables are called the *dependent* variables. In the case described, the dependent variable was the average number of words correctly recognized for each type size.

It is not necessary, in an experiment, that the independent variable be one which can be varied quantitatively. It merely happened to be so in the example cited. Many experiments are concerned, for example, with a comparison between what we have previously called various treatments. The questions asked in these experiments have to do with differences in the dependent variable under different treatments. Experiments of this kind have been described as *comparative* experiments. In a comparative experiment interest is directed toward the problem of discovering whether the different treatments result in differences in the observed values of the dependent variables. When the treatments represent a quantitative series, then we may also be interested in studying the functional relationship between the quantitative independent variable and the dependent variable, that is, in determining whether the relationship is linear or of some other form.

The advantages of making observations under controlled conditions over observations without such control have been pointed out by Woodworth (1938, p. 2):

1. The experimenter makes the event happen at a certain time and place and so is fully *prepared* to make an accurate observation.

2. Controlled conditions being *known* conditions, the experimenter can set up his experiment a second time and repeat the observation; and—what

is very important in view of the social nature of scientific investigation—he can report his conditions so that another experimenter can duplicate them and check the data.

3. The experimenter can systematically *vary* the conditions and note the concomitant variation in the results.

QUESTIONS AND PROBLEMS

1. Can you think of a case in the behavioral sciences in which incidental observation was instrumental in the formulation of some hypothesis which was then investigated by means of systematic observation?

2. Suppose someone is interested in IQ's of 1,000 fifth-grade children, that is, he has available 1,000 such observations. Of what value would some form of data reduction be in this instance?

3. Make a list of 5 response variables, 5 stimulus variables, and 5 organismic variables other than the ones mentioned in the chapter. What available methods are there for quantifying each of the variables? How might those variables for which methods are not available be quantified?

4. Assume that one of the variables in a research problem is socio-economic status. If you were constructing an index or a test for this variable, what factors, in addition to income, would you want to take into consideration?

5. A fortunate basketball coach at a small college once had 5 players trying for the position of center on the college team. The members of the coaching staff were unable to differentiate between the abilities of the 5 players. What situation tests might be developed to yield quantitative data concerning the ability of each player for the position of center?

6. A graduate department of psychology awards a number of research fellowships in psychology to students who are believed to be outstanding. Assume that the awards are to be based primarily upon the potentiality of the students to do research in the field of psychology. What factors should be taken into consideration in making the awards? What methods might be devised for quantifying the variable of interest?

7. Comment upon the following statement: "If a variable is truly continuous, then, in theory, no two observations could ever have the same value."

8. What justification can be given for treating scores on psychological tests as continuous measurements? The point of view taken in this text is elaborated on by Edwards (1958). For a different point of view, see Stevens (1951) and Siegel (1957).

9. Select and read a research article in some journal. What question or questions was the research attempting to answer? What variables were involved in the research? What was the nature of the observations made?

10. In psychoanalytic theory, the *id* is said to be that aspect of the individual concerned with instinctual reactions for satisfying motives. According to Morgan (1956, p. 633), "The id seeks immediate gratification of motives with little regard for the consequences or for the realities of life." Regardless of whether or not this characteristic is called the id or something else, it seems reasonable that

individuals do differ in the degree to which they manifest the characteristic. What observations might be made to obtain some measure of the "strength" of the id?

11. Psychologists who are interested in personality research make use of a large number of variables that refer to personality traits or characteristics (organismic variables?). Some examples are honesty, ego control, dependency, achievement motive, deference, and social introversion. What are some of the most frequently used methods of observing these variables?

12. Comment upon the following statement: "Naming a variable may suggest the observations to be made, but, in a very real sense, the observations actually made define the variable itself."

13. Define, briefly, each of the following terms:

behavioral variable
comparative experiment
continuous variable
dependent variable
discrete variable
enumeration data
experiment
frequency data
independent variable
organismic variable
qualitative variable
quantitative variable
response-inferred organismic variable
stimulus variable
treatment
unordered variable
variable

2

PRINCIPLES OF EXPERIMENTAL DESIGN

INTRODUCTION

The set of observations that one makes in doing research on a question of interest is called a *sample*. A sample of observations is but a portion of the complete set of all possible observations relevant to the same question. This larger group of potential observations is called a *population*. A population does not necessarily refer to individual persons. A population might consist of the observations of height or weight or of other organismic variables that we assume to be characteristic of individuals. But it might also consist of the observations that one might make with respect to all schools in a given city or all third-grade classes in a given school.

Some populations can be described as *finite*, that is, the number of observations that can be made is limited. The observations of age of all faculty members at a given university, the observations of intelligence test scores of all students registered in introductory psychology at the same university, and observations of the number of letters in each word in a given text are all examples of finite populations. For finite populations, the total number of possible observations can, at least in theory, be enumerated or listed.

Still other populations can be described as *infinite*. For infinite populations the total number of possible observations cannot be enumerated. For example, if a variable of interest is the face of a coin after it has been tossed, the observation we might make after each toss is whether the coin has fallen heads or tails. We can conceive of the coin being tossed an indefinitely large number of times so that we can make an indefinitely large number of observations. In this instance, the number of potential observations that we could make is unlimited and the population is described as infinite.

In general, when a sample of observations is used as a basis for answering a question, we do not wish the answer to be confined or restricted to the particular sample of observations made. Instead, we wish to obtain an answer such that we have some degree of confidence the answer is also

pertinent to the population from which the sample is drawn. In fact, many, if not all, of the questions we ask in research and experimentation are motivated not by interest in a particular sample but rather by our interest in the population. In other words, we want to use the sample of observations to arrive at an answer to a question concerning the population. The process of using a sample to infer something about a population is known as *statistical inference.*

As we have indicated earlier, the techniques used in statistical inference make the evaluation of a given set of observations objective, but the inference made may be right or wrong. This is a situation which must always be faced whenever we use a sample as a basis for inferring something about a population. In this chapter we shall describe two simple experimental designs and illustrate how, in each case, the observations made may be evaluated by means of statistical methods. In our discussion of these two designs, we can examine some of the problems involved in statistical inference. In the following chapter, we shall give further consideration to these problems and describe the particular populations relevant to each of the two designs.

THE FARMER FROM WHIDBEY ISLAND

A farmer from nearby Whidbey Island visited the psychological laboratory of the University of Washington. He had with him a carved whalebone and claimed that in his hands the bone was an extremely powerful instrument capable of detecting the existence of even small quantities of water. To support his claim he said that several of his neighbors on Whidbey Island had tried unsuccessfully to bring in water wells. Finally they had called upon him for help. He had taken his whalebone, grasped one fork in each hand, and walked slowly over their ground. Suddenly, the point or apex of the bone had dipped sharply toward the ground. When his neighbors had drilled wells at the points he had located in this fashion, they had found water.

The farmer added that he was unable to explain his peculiar power. His neighbors were unable to use the whalebone in locating water. It had to be in his hands before it would dip sharply, indicating the presence of water. He was somewhat disturbed by his ability, and he thought that perhaps the psychologists at the university would be interested in examining him and telling him why it was that he was able to use the bone so effectively while others could not. He himself thought that it had something to do with "magnetism" that emanated from his body. Anyway, he would be willing to demonstrate his ability so that the psychologists could see for themselves. Perhaps then they could explain it to him.

At this point in his story, the farmer asked that he be given a paper cup filled with water. When he was given the cup, he placed it on the floor. He then grasped the whalebone and held it stiffly in front of him as he moved slowly about the room. When the apex of the bone passed over the cup of water, his arms trembled slightly and the bone dipped toward the ground. The farmer showed signs of strain and remarked that the force was so powerful he was almost unable to keep the bone in his grip.

The psychologist thanked the farmer for his demonstration and said that he would like to test the farmer's ability to locate water under controlled conditions, but that this would require some preparation. Would the farmer agree to return for these tests next week? The farmer agreed and promised to return at the appointed time.

THE FIRST EXPERIMENT

When the farmer returned to the psychological laboratory the next week, he was greeted by the psychologist and taken to one of the laboratory rooms. Spread around the floor of the room were 10 pieces of plywood about 8×8 inches in size. Numbers from 1 to 10 had been marked upon the top of each square. The pieces of plywood were resting upon tin cans, with the labels removed, about No. 2 in size. The psychologist explained that 5 of the cans had been filled with water and that 5 had been left empty. He had not used any systematic basis in determining which 5 cans were to be filled with water and which 5 were to be left empty, but rather, as he put it, "this was left to chance." As a matter of fact, he added, he himself did not know which of the cans contained water and which were empty, since he had left this task to a laboratory assistant. He was as much in the dark as the farmer, but he hoped that the farmer, with the aid of his whalebone, would soon be able to enlighten him. He again emphasized to the farmer that under 5 of the sections of plywood were cans with water and under 5 other sections the cans were empty, and that the arrangement of the empty and filled cans was purely a chance or random one.

The psychologist now wanted the farmer to take his whalebone and attempt to divide the 10 squares of plywood into two groups. One group would be the 5 squares covering the cans filled with water and the other group would be the 5 squares covering the empty cans. The farmer did not need to make his choice in any particular order; he was merely to divide the set of 10 sections of plywood into two groups of 5 each.

The observations to be made in this experiment consist of the choices that the farmer makes. The outcome of the experiment is the particular set of choices that the farmer does make. We shall examine this experiment in some detail. We shall pay particular attention to possible outcomes of the experiment, the question which the experimenter hopes to answer by the

observations made, and the manner in which he proposes to arrive at this answer.

The Question of Interest

The question which motivates the experimenter to make the observations is not necessarily the one of interest to the farmer. The farmer, in his previous conversation, had indicated that he wanted to know why he could divine the presence of water. It is apparent that the farmer implicitly assumes that he can detect the presence of water. The question of interest to the psychologist, on the other hand, is whether or not the farmer is successful in doing what he believes he can do; that is, can he actually detect the presence of water?[1]

The psychologist may reason in this way: Let us assume that the farmer does *not* possess any particular powers which enable him to locate water with his whalebone; that the only factor which is operating in determining his choice is chance. More specifically, the question which the experimenter wishes to answer is: Can the farmer do any better in his choices than might be expected on the basis of chance?

Permutations

The possible outcomes of this experiment can be demonstrated in a simple way by the rules for permutations and combinations. *Permutations* refer to the number of arrangements (orders) in which a set of n distinct objects may be arranged. In general, the number of permutations of n distinct objects is given by

$$_nP_n = n! \tag{2.1}$$

where $n!$ is called factorial n and represents $(n)(n-1)(n-2)\cdots(2)(1)$, or the product of all of the successive integers from n to 1. Factorial 0! is always taken equal to 1. The number of permutations of n objects taken r at a time is given by

$$_nP_r = \frac{n!}{(n-r)!} \tag{2.2}$$

In the problem at hand, the number of orders in which 5 sections of plywood may be selected from the available 10 is

$$_{10}P_5 = \frac{10!}{(10-5)!} = 30{,}240$$

This figure gives every possible set of 5 things arranged in every possible order, that is, any one of the 10 sections may be selected first; this choice

[1] To ask why the farmer is successful in divining the presence of water is meaningful only if it can first be demonstrated that he is, in fact, successful.

may be followed by any one of the remaining 9 things; this choice may be followed by any one of the remaining 8; and so on until five have been selected.

Combinations

But in this experiment the psychologist is not going to demand that the farmer select the set of 5 cans containing water in any particular order. All that the psychologist is interested in is the set of 5, once the set has been selected. As far as he is concerned, the set of cans 10, 5, 8, 2, and 3, selected in that order, is equivalent to the set of cans 8, 3, 2, 10, and 5, selected in that order, or in any other possible order.

It may be noted that the set of 5 selected objects or sections may themselves be arranged in (5)(4)(3)(2)(1) = 120 orders, according to formula (2.1). Thus, dividing 30,240 by 120, we obtain 252 ways in which a set of 5 objects may be selected from 10, if the *arrangement or order is ignored.* In general, the number of *combinations* (arrangement ignored) of n distinct objects taken r at a time is given by

$$ {}_nC_r = \frac{{}_nP_r}{{}_rP_r} = \frac{\dfrac{n!}{(n-r)!}}{r!} = \frac{n!}{r!(n-r)!} \tag{2.3} $$

or, in the present problem,

$$ {}_{10}C_5 = \frac{10!}{5!(10-5)!} = \frac{(10)(9)(8)(7)(6)}{(5)(4)(3)(2)(1)} = 252 $$

Test of Significance

Now the best that the farmer could possibly do in the present experiment would be to select the particular set of 5 which happened to be those with water in the cans. There is only one way in which this could happen and this particular selection would be 1 out of 252 possibilities. If only chance factors are operating in determining the selection and if this experiment with the farmer were repeated an indefinitely large number of times, then we would expect this particular set to be selected with a theoretical relative frequency of 1/252.

By probability, symbolized by P, we shall mean a theoretical relative frequency. Thus, we have $P = 1/252 = .004$ (more precisely, .00397). We may say that this result, 5 correct choices, would be expected by chance alone only about 4 times in 1,000.

In essence, we have made a *test of significance.* We started by assuming that the farmer would respond to the test situation on the basis of chance. This assumption may be regarded as a hypothesis, often referred to as a *null hypothesis,* which the experiment is designed to test. On the basis of

this hypothesis it is possible to determine the theoretical relative frequency of 5 correct choices, *assuming the hypothesis to be true.* The probability of .004 is the final result of our test of significance. If the probability yielded by the test of significance is small, then either the hypothesis and its related assumptions are false, or else an unusual, that is, a rare or improbable, event has occurred. The test of significance, resulting in a probability, is simply a method of enabling the experimenter to determine whether or not he wishes to regard the hypothesis being tested, and its related assumptions, as tenable or untenable.

SIGNIFICANCE LEVELS

The probability corresponding to the occurrence of any given event or events may range in value from 0 to 1. If the probability is 1, then the event is certain to happen. If the probability is 0, then the event is an impossible one. A probability of .95 refers to an event that may be expected to occur 95 times in 100, while a probability of .05 refers to an event that may be expected to happen 5 times in 100. In other words, the larger the probability, the more likely the event is to happen, and the smaller the probability, the less likely the event is to happen.

In testing hypotheses, we must decide how small the probability of a given event must be before we shall choose to regard the event as improbable. There are many aspects to this question and the choice is best determined by considering the particular problem under investigation. We shall, however, need some guidepost for subsequent discussions and, for this purpose only, we shall choose a value that is frequently used by research workers. We shall regard as a small probability one that is equal to or less than .05. If, in evaluating the outcome of a given experiment, our test of significance results in a probability of .05 or less, assuming the null hypothesis to be true, then the outcome we have obtained is one that would occur 5 times or less in 100. When this occurs, we shall reject the null hypothesis tested. The probability that we choose to use in rejecting the null hypothesis is called the *significance level* of the test and is symbolized by α. If we choose α equal to .05 and our test of significance results in a probability of .05 or less, we say the result is significant at the 5 per cent level.

We emphasize that the significance level of a test should be chosen in advance of making the test, and that there are many other values of α rather than .05 which could be used as the significance level. But it is apparent that if we choose α very small, we decrease the probability of rejecting the null hypothesis. Carrying out an experiment would be useless if the experimenter refused to reject the hypothesis tested regardless of how improbable it is in terms of the results he obtains.

It should be made clear also that no single experiment can establish the *absolute proof* of any conclusion, however significant (regardless of the smallness of the probability) the result of the test of significance may happen to be. The 1 chance in 100 ($P = .01$), the 1 chance in 1,000 ($P = .001$), or the 1 chance in 1,000,000 ($P = .000001$), for that matter, "will undoubtedly occur, with no less and no more than its appropriate frequency, however surprised we may be that it should occur to us. In order to assert that a natural phenomenon is experimentally demonstrable we need, not an isolated record, but a reliable method of procedure. In relation to the test of significance, we may say that a phenomenon is experimentally demonstrable when we know how to conduct an experiment which will rarely fail to give us a statistically significant result" (Fisher, 1942, pp. 13–14).

TWO TYPES OF ERRORS

As we have indicated above, in making tests of significance we shall sometimes be in error in the inference drawn concerning the hypothesis tested. When the null hypothesis is true and the results of our test of significance reject it, or declare it false, we describe this as a *Type I error.* When the null hypothesis is false and the results of our test of significance fail to reject it, or fail to declare it false, we describe this as a *Type II error.* The probability of making a Type I error is set by α, the significance level we have chosen. If we always reject a hypothesis when the test of significance yields a probability of .05 or less, and if we consistently follow this standard, then we shall incorrectly reject 5 per cent of the true hypotheses tested, that is, we shall declare them false when they are in fact true. If we demand that the probability be .01 or less before rejecting a hypothesis, and if we consistently follow this standard, then we shall incorrectly reject as false 1 per cent of the true hypotheses tested. By choosing α small we decrease the probability of a Type I error. But, at the same time, we increase the probability of a Type II error.

In general, if we hold α, the significance level, constant, we can decrease the probability of a Type II error by increasing the number of observations in our sample. We shall have more to say about these two types of errors in later discussions.

EXPERIMENTAL CONTROLS

In the experiment described, if the farmer makes 5 correct choices, we know that the probability of this outcome, under the null hypothesis that his choices are a matter of chance, is .004. Hence, with a significance level of .05, if the farmer is able to choose the particular set of 5 cans with water

with the aid of his whalebone, then we should conclude that the probability of this occurring by chance alone is so small that the hypothesis should be rejected.

At this point we shall do well to consider what the rejection of the hypothesis tested means. If the hypothesis is rejected, this means only that the experimenter is not willing to assume that chance alone determined the farmer's choices. It does not prove that the whalebone had any particular influence upon the farmer's choice. The test of the null hypothesis had nothing to do with *why* the farmer was able to choose correctly. The psychologist might be willing to *assume* or *infer* that the whalebone played some part in the farmer's selections, but he would undoubtedly do this only if other possible explanations had been ruled out in terms of *experimental controls*. What are some of these alternative explanations?

Without the experimenter knowing about it, the farmer may use the toe of his foot to tap the cans under the board. Since, in this manner, the cans filled with water could be easily distinguished from the empty cans, this alone would account for a perfect selection upon the part of the farmer. If this is the basis of the farmer's selections, then obviously the whalebone has nothing to do with his choices. The farmer might even deny that he is using this cue—the sound of the can when tapped with his foot—if questioned about it. But the psychologist knows that many of our choices and judgments are based upon factors of which we are not aware. It would be the experimenter's responsibility to rule this possibility out by observation, or by some other control.

Again the psychologist would want to make sure that the farmer does not tap the tip of the whalebone on the tops of the plywood sections. If the farmer does this, his choice might be determined by the differences in sound of the sections covering the empty and filled cans. He might thus make a perfect selection and the experimenter would reject the hypothesis of chance. But note again the rejection of the hypothesis of chance does not establish the validity of the farmer's claim concerning the influence of the whalebone.

Another possible explanation of a perfect selection might be that the experimenter's assistant had spilled some of the water on the floor in filling the cans. The water might have been carefully mopped up, but slight cues may have remained. The absence of dust or the cleanliness of the floor under the sections of plywood containing water, as a result of the mopping, might provide cues for the farmer's choice.

If the experimenter, rather than his assistant, had filled the cans, so that he had knowledge of which cans contained water and which did not, then the experimenter himself might give some sign: a holding of his breath, a biting of his lips, or some other unconscious gesture, as the farmer moved his whalebone over the sections containing water. The farmer's choice might

thus be based upon one of the unconscious gestures or reactions of the experimenter, without, of course, the experimenter, and perhaps even the farmer, being conscious of the fact that these cues were the basis of the farmer's choice. Fortunately, the experimenter, in this instance, anticipated this possibility and controlled it by having his assistant prepare the cans.

In a well-designed experiment, the various factors which may influence the outcome of the experiment and which are not themselves of interest must be controlled if sound conclusions are to be drawn concerning the results of the experiment. It is to be emphasized that these conclusions are derived from the structure of the experiment and the nature of the controls exercised. They do not come from the test of the null hypothesis. The statistical test indicates only the probability of a particular result upon the basis of the statistical hypothesis tested, namely, in the case described, that chance alone is determining the outcome. If the experimenter rejects the hypothesis of chance, he must still examine the structure of his experiment and the nature of his experimental controls in making whatever explanation he does make as to why he obtained the particular result he did.

THE IMPORTANCE OF RANDOMIZATION

An essential notion in evaluating the outcome of the experiment described is that of *randomness*. We may recall that the experimenter mentioned that the selection of the 5 cans to be filled with water and the 5 to be left empty was determined on a random basis. The randomization in the assignment of the treatments, in this instance, served two functions. For one thing, since the randomization was done by the assistant, the experimenter, who made the observations of the farmer's choices, was in ignorance as to which cans contained water and which were empty. Thus, the randomization of the treatments offers assurance that the experimenter himself would not provide cues which would assist the farmer in making correct selections.

Randomization also insures that the particular model used in evaluating the results of an experiment is applicable. Suppose, for example, in the experiment described, that some slight but perceptible differences existed in the cans such that 4 of the cans had a slight dent. If the assistant had systematically, but not necessarily consciously, filled either the dented cans or 5 of the undented cans, the farmer may have reacted to this cue and used it as a basis for his choices. If the assistant had observed the dents in the 4 cans, he may have controlled for this factor by selecting, at random, 2 of the dented cans to be filled with water and 2 to be left empty. But suppose also that among the set of 10 cans, 6 cans have a trace of rust and 4 cans have not. Again, if any division between the filled and unfilled cans resulted in an association with this characteristic, this may also provide the

farmer with a basis for making his choices. Since there are any number of possible characteristics of the cans which might be associated with a systematic division of the set of 10 into two groups of 5 each, the only satisfactory basis for dividing the set is that of randomization.

Similarly, in assigning the numbered sections of plywood to the 10 cans, randomization is again necessary. For, in this instance, the assistant might assign the filled cans the even numbers and the empty cans the odd numbered sections. Randomization, at this stage, is necessary in order to insure that there is no association between the characteristics of the pieces of plywood and the numbers on them, on the one hand, and the filled and empty cans themselves, on the other hand.

If the experimenter had made a systematic division of the cans into two sets, he may have done so on the assumption that the manner in which the division was made could not in any way influence the observations to be made in the experiment, that is, the farmer's choices. This assumption may, of course, be true, but it remains an assumption. Even though the experimenter may be convinced of the truth of the assumption he has made, he may have difficulty in convincing others. The only really convincing argument is that of appropriate randomization. We shall have much more to say about the importance of randomization in later discussions.

A LIMITATION IN THE FIRST EXPERIMENT

Let us suppose that, in the experiment described, the farmer claims that the psychologist has set too high a standard; that occasionally the whalebone fails him and that the psychologist should not expect him to make a perfect selection of the 5 cans. What will the attitude of the psychologist be, for example, if the farmer selects 4 cans with water, but for his fifth choice makes an error and selects 1 of the 5 empty cans?

The experimenter's attitude will again depend upon α, the level of significance, and upon the probability of this result, assuming the null hypothesis to be true. To evaluate this result, we first find the number of ways in which it can occur. Ignoring the order of selection of the cans, we may note that from the set of 5 cans with water, 4 cans may be selected in 5 ways. This is given by formula (2.3). Thus,

$$ {}_5C_4 = \frac{5!}{4!(5-4)!} = 5 $$

Independently of this selection, 1 can may be selected from the set of 5 dry cans in 5 ways also, as determined by the formula for combinations. Hence, there are $(5)(5) = 25$ ways in which this particular event can occur.

The probability that the farmer will select 4 cans with water and 1 empty can, by chance, will then be 25/252 or approximately .099. In

evaluating the null hypothesis, however, we need to consider not only the probability (.099) of 4 wet cans and 1 dry can being selected, but also the probability (.004) that 5 wet cans are selected. Both of these outcomes would offer evidence against the null hypothesis. Since the two outcomes are mutually exclusive, the probability of the farmer making 4 *or more* correct choices will be equal to .099 + .004 = .103.

It is clear, if $\alpha = .05$, that within the scope of this experiment only a perfect selection by the farmer would result in the rejection of the null hypothesis This particular experiment permits no possibility of an error upon the part of the farmer and, in this respect, may be considered too demanding.

THE SECOND EXPERIMENT

We shall examine briefly a variation in the experimental procedure that would permit the farmer to make an error and still permit the rejection of the null hypothesis. In this variation 10 additional cans are obtained, and the 20 cans are arranged at random in 10 pairs. One member of each pair is filled with water and the one to be filled is again determined at random. The farmer is told that he will be presented with 10 pairs of cans, one of which is filled with water and one of which is empty, and that he is to select the member of each pair that he believes contains the water. What are the possible outcomes of this experiment, again assuming that in each pair the farmer's choice will be based upon chance alone?

There are two ways in which the farmer's first choice may be made, and, independently of this choice, the second choice may be made in two ways, and, independently of this choice, there are two ways in which the third choice may be made, and so on for the 10 choices. Thus, there are a total of 2^{10} or 1,024 ways in which the farmer may make his selections. Each choice may be judged "right" if the can containing water is selected and "wrong" if the empty can is selected, so that the possible outcomes may be recorded as 10, 9, 8, 7, 6, 5, 4, 3, 2, 1, and 0, right.

There is only one way in which 10 right choices may be made, that is, the water-filled can would have to be selected in every pair of the 10 pairs presented, and this can occur in only one way. A selection of 9 right and 1 wrong can be made in 10 ways. The number of ways in which each of the other possible results may occur can be obtained by means of formula (2.4). This formula gives the number of permutations of n objects when these can be divided into k sets so that the objects within each set are alike. We let $r_1, r_2, \cdots, r_k$ represent the number of objects in each of the respective sets, with $n = r_1 + r_2 + \cdots + r_k$. Then

$$ {}_nP_{r_1, r_2, \cdots, r_k} = \frac{n!}{r_1!\, r_2! \cdots r_k!} \tag{2.4} $$

For the case at hand, we have only two sets, the number of right and the number of wrong choices, with n equal to 10. Thus, for the number of ways in which we may have 10 right and 0 wrong, 9 right and 1 wrong, 8 right and 2 wrong, and so on, we have

$$
\begin{aligned}
{}_{10}P_{10,0} &= \frac{10!}{10!\,0!} = 1 \\
{}_{10}P_{9,1} &= \frac{10!}{9!\,1!} = 10 \\
{}_{10}P_{8,2} &= \frac{10!}{8!\,2!} = 45 \\
{}_{10}P_{7,3} &= \frac{10!}{7!\,3!} = 120 \\
&\vdots \qquad\quad \vdots \qquad\quad \vdots \\
{}_{10}P_{0,10} &= \frac{10!}{0!\,10!} = 1
\end{aligned}
$$

With repetitions of the experiment, therefore, we would expect a set of 10 right and 0 wrong choices to occur with a theoretical relative frequency of 1/1,024, a set of 9 right and 1 wrong choices with a theoretical relative frequency of 10/1,024, a set of 8 right and 2 wrong choices with a theoretical relative frequency of 45/1,024, and so on, if nothing but chance is determining the choices. The value of P for 10 correct choices is thus 1/1,024 or approximately .001 (more precisely, .00098) and this outcome can be expected to occur about 1 time in 1,000. If $\alpha = .05$, then this result would be regarded as significant and the null hypothesis would be rejected.

If the farmer makes 9 correct choices, we find that the probability of this occurring is 10/1,024, yielding a P of approximately .01 (more precisely, .00977). In evaluating this outcome, we again need to consider any result more extreme which also offers evidence against the null hypothesis, that is, the probability of 10 correct choices. Thus, the probability of 9 *or more* correct choices is given by the sum of the ways in which 9 and 10 correct choices may be made, divided by the total number of ways. We find, therefore, that the probability of 9 or more correct choices is 11/1,024 or .011 (more precisely, .01074). Similarly, the probability of 8 *or more* correct choices, by chance, is given by the sum of the ways in which 8, 9, and 10 choices may be made, divided by the total number of ways, or 56/1,024, which yields a P of .055 (more precisely, .05469). Seven or more correct choices would result in a probability of 176/1,024 or .172 (more precisely, .17187). Seven or 8 correct selections upon the part of the farmer would

thus not offer significant evidence against the null hypothesis, with α equal to .05, whereas 9 or 10 correct choices would.

It may be noted that this particular experimental design permits the farmer to make at least one error, something that the first experimental design did not, and still enables the experimenter to reject the null hypothesis. However, even this experiment is also fairly limited in the sense that the farmer is required to be correct in at least 9 out of 10 choices before the experimenter is willing to conclude that he is not responding by chance.

With adequate experimental controls and the appropriate use of randomization, we may be confident that the farmer is not responding by chance, if he makes 9 or 10 correct choices, that is, if we get a significant outcome. On the other hand, if we fail to get a significant result, that is, if we fail to reject the null hypothesis, we should be aware that we may be making a Type II error. We have not explored this possibility in any detail. The farmer, for example, *may* be able to do better than chance. This particular experiment, as it stands, however, is not very sensitive to detecting an ability upon his part that enables him to do only *slightly* better than chance. It could be made more sensitive, more likely to detect choices only slightly better than chance, by increasing the number of observations. In later discussions we shall see why this is so.

QUESTIONS AND PROBLEMS

1. A rat is placed upon a Lashley-type jumping stand and through a series of trials is trained to jump always to the smaller of two squares. The right and left position of the smaller square is randomly alternated so that the experimenter has some confidence that the rat is not reacting to a position variable. The experimenter is interested in determining whether the established reaction pattern will be generalized to the extent that the rat will react similarly to the smaller of two circles. After the rat has learned to discriminate between the two squares, it is given a series of 8 trials with two circles. The position of the smaller of the two circles is randomly alternated. We make the assumption that if generalization of the previous learning is not present, the rat will react to the two circles on the basis of chance. On the other hand, if the rat jumps to the smaller circle with a frequency greater than we are willing to attribute to chance, this hypothesis will be ruled out and we shall infer that generalization has taken place. (*a*) What is the probability of 7 or more jumps to the smaller circle in 8 trials if the null hypothesis is true? (*b*) If the number of trials is increased to 12, what is the probability of 10 or more jumps to the smaller circle if the null hypothesis is true?

2. It is claimed that infants stimulated by a loud sound show a response pattern that is differentiated from the pattern of response present when movements are restrained. Response to the loud sound is said to be that of "fear" and response to restraint is said to be that of "rage." An infant is stimulated 4 times by loud sound and 4 times by restraint of movement, and motion pictures are taken of the responses immediately after stimulation. Photographs are made from

the film and printed in strips. There are a total of 8 strips, 4 showing reaction to sound and the other 4 to restraint. This is explained to subjects who are to serve as judges. They are asked to select the set of 4 showing "fear." The questions which follow are related to the evaluation of the possible outcomes of the experiment, under the hypothesis that a correct selection is a matter of chance. (*a*) What is the probability of a single subject selecting the set of 4 correct photographs from the 8? (*b*) What is the probability of selecting a set of 3 correct and 1 wrong? (*c*) What is the probability of selecting a set of 2 correct and 2 wrong? (*d*) Suppose that the experiment had made use of 12 photographs, 6 of rage and 6 of fear. What are the possible outcomes of a subject's judgments and what is the probability of each?

3. An experimenter has a set of 4 cards, 3 of which are blank and 1 of which has an X printed on it. The cards are shuffled and placed face down in a row. The subject is to determine the position of the card with the X on it. (*a*) What is the probability that the subject will make a correct selection in a single trial, assuming that he is reacting by chance? (*b*) If the subject is given 4 trials, what is the probability that he will make precisely 3 correct choices by chance? (*c*) What is the probability that he will make precisely 1 right choice in 3 trials? (*d*) If there are 128 subjects who serve in the experiment and each subject is given 3 trials, then how many subjects would be expected to obtain perfect scores by chance alone? (*e*) How many of the subjects would be expected to obtain scores of 2 or more correct by chance?

4. An investigator has given a battery of 15 tests to a group of students. If he should be interested in the relationship between each test with every other test, how many such relationships would he have to study?

5. A student claims that he can differentiate his own brand of cigarettes from three other popular brands. Outline an experiment which would provide evidence with respect to this claim. What kinds of experimental controls might be necessary in the experiment?

6. Outline an experiment in which one might test a student's claim that he can discriminate Beer A from Beer B. What experimental controls may be necessary? What role does randomization play in the experimental design? What outcomes of the experiment would be regarded as significant?

7. Suppose we wish to determine whether orange juice can be distinguished from onion juice and apple juice when visual and olfactory cues have been experimentally controlled. We block the nasal passages of a subject and blindfold him. He is then presented with a set of 3 test tubes. He is told that one of the test tubes contains onion juice, one orange juice, and one apple juice and that he is to pick out the one which he thinks contains the orange juice. Fifteen sets of 3 test tubes are presented to the subject. (*a*) What is the probability that he will make 9 or more correct choices, if he is responding by chance? (*b*) What are some of the experimental controls that should be considered in planning this experiment? For example, what about the temperature of the juices?

8. In many experiments, the dependent variable is a rating assigned to a subject or an object by a judge. For example, judges may be asked to rate the improvement of patients in a mental hospital after they have been treated with a drug or after they have had several months of psychotherapy. In taste labo-

ratories, judges may be asked to rate the quality of foods that have been differently prepared. Discuss the nature of the experimental controls necessary in such studies. Discuss the role of randomization as a device for concealing from the judges which patients have been treated and which have not, and similarly which foods have been prepared in one way and which in another.

9. Define, briefly, each of the following terms:

alpha	probability
combinations	sample
experimental controls	significance level
finite population	statistical inference
infinite population	test of significance
null hypothesis	Type I error
permutations	Type II error

3

BINOMIAL POPULATIONS IN RESEARCH

INTRODUCTION

In the first experiment with the farmer from Whidbey Island, the task assigned to him was to divide 10 cans into two sets, one set containing the 5 cans filled with water, and the other set the 5 empty cans. If we know the nature of the observations in one of these two sets, then we also know the nature of the observations in the other. Consider, for example, only the set of 5 which the farmer says contain water. If 4 of his choices in this set are right, that is, are cans containing water, and 1 is wrong, then we know that the other set must have 4 empty cans and 1 with water. We may confine our attention, therefore, to only one of the two sets, since no additional information will be provided by considering both. Let us consider the set of 5 that the farmer says contain water. We shall regard this set as a *sample* of 5 observations from a *finite population* of 10.

RANDOM SELECTION AND RANDOM SAMPLES

The finite population can be simulated by placing 10 disks in a box. We shall let each disk correspond to an observation without, for the moment, specifying the value of the observation. It will be convenient, however, if we identify the disks in the same manner in which the 10 pieces of plywood were identified, that is, by the numbers from 1 to 10. We now shake the box thoroughly and then let one disk fall through a small slot in the box. We shall assume that this procedure results in a random selection of a disk. By *random selection*, we mean that we shall assume that the probability of any given disk in the box falling through the slot is the same for all disks. Having selected one disk, *without replacing it in the box*, we again shake the box and draw a second disk from the remaining 9. We again shake the box and draw a third disk from the remaining 8. We continue in this manner until we have a sample of 5 disks or observations. We let this sample correspond to a set of 5 choices that could be made by the farmer in the experiment.

If our method of sampling is random, then on each draw each of the disks in the box has an equal probability of being selected. For example, the probability of a particular one of the 10 disks being selected on the first draw is $\frac{1}{10}$; on the second draw, with 9 disks in the box, the probability of a particular one being drawn is $\frac{1}{9}$, and so on, until on the last draw the probability of a particular one being drawn is $\frac{1}{6}$.

What is the probability that a sample, drawn in the manner described, will include the observations identified by the numbers 10, 8, 5, 4, and 1? Consider first the probability of obtaining these 5 observations in the order specified. The probability of obtaining 10 on the first draw is $\frac{1}{10}$. Given that we have drawn 10, the probability of 8 on the second draw is $\frac{1}{9}$. Given that we have obtained 10 on the first draw and 8 on the second, the probability of 5 on the third draw is $\frac{1}{8}$, and so on. The probability that the sample will be the particular set of 5 observations drawn in the order specified is

$$\frac{1}{10} \times \frac{1}{9} \times \frac{1}{8} \times \frac{1}{7} \times \frac{1}{6} = \frac{1}{30{,}240}$$

If we are not interested in the order in which these particular 5 observations are drawn, but simply in the probability that the sample will contain the specified 5 observations, then we note that the observations may be permuted in $5! = 120$ ways. The probability that a sample selected in the manner described will contain the specified 5 observations, the order in which they are drawn being immaterial, will be $120/30{,}240 = 1/252$.

The probability that we have just obtained will be exactly the same for any other specified set of 5 observations differing from the sample considered in one or more observations. For example, the probability that the sample will contain the observations 10, 4, 3, 2, and 1 is also 1/252. The total number of different samples will be given by the number of ways in which 5 disks can be selected from 10, with the order of selection ignored, and, by means of formula (2.3), we see that this is

$$\frac{10!}{5!(10-5)!} = 252$$

We have just seen that every possible different sample has a probability of 1/252 of being drawn and the probability that we will obtain one of the 252 possible samples is, of course, 1.00.

The probability that any specified observation will be included in the sample of 5 observations is $\frac{1}{2}$. To see that this is so, we consider the probability that the specified observation will be the first one drawn. This probability is $\frac{1}{10}$. The probability that the observation will be the second one drawn will be the product of the probability that it is *not* drawn first times the probability that it will be selected from the remaining 9. Thus $(\frac{9}{10})$

$(\frac{1}{9}) = \frac{1}{10}$. Similarly, the probability that the observation will be the third one selected will be given by $(\frac{9}{10})(\frac{8}{9})(\frac{1}{8}) = \frac{1}{10}$ In the same manner, we find that the probability of the observation being the fourth one drawn is $\frac{1}{10}$ and this is also the probability that it will be the fifth one drawn. Since these are mutually exclusive events, by the addition rule, the probability that the observation will be included in the sample is $\frac{5}{10} = \frac{1}{2}$. This probability is the same for each of the 10 observations.

We can thus say that every possible sample of 5 observations has the same probability (1/252) of being drawn and that every observation has the same probability ($\frac{1}{2}$) of being included in the set of 5 observations. These properties are often used to define a *simple random sample* or, more briefly, a random sample. The use of the term random in connection with a sample of observations should be considered as applying, however, to the particular procedure or method of selecting the observations, rather than to the sample itself. A random sample of observations, in other words, is one obtained by a particular method which we believe introduces randomness into the selection of the observations. In the present instance, we assumed that randomness in the selection of the observations was introduced by the use of our sampling box in which the disks were thoroughly mixed before one was selected. More useful procedures of random selection will be discussed later.

PROBABILITIES OF POSSIBLE OUTCOMES

Let us now assume that we have assigned, again by random methods, a value to each of the 10 disks in such a way that 5 of the disks or observations have a value of W, corresponding to a filled or wet can, and 5 have a value of D, corresponding to an empty or dry can. Our method of sampling remains the same, but we are now interested in the number of W's in each of the 252 possible samples of 5 observations each. We already know that only one of the 252 possible samples can include the 5 observations with values of W and we can say that the probability of obtaining a sample of 5 W's is, therefore, 1/252.

What is the probability of obtaining a sample of 4 W's and 1 D? Specifically, let the first 4 observations drawn have the value of W and the last one the value of D. The probability of obtaining an observation with a value of W on the first draw is $\frac{5}{10}$; if this occurs, then there will be 4 observations with W's left in the box and 5 with D's and the probability of a W on the second draw will be $\frac{4}{9}$. If the first two draws are W's, then there will be 3 W's and 5 D's left in the box and the probability of W on the third draw will be $\frac{3}{8}$. If we obtain a W on the third draw, then there will be 2 W's left and 5 D's. The probability of W on the fourth draw will then be $\frac{2}{7}$. If this occurs then we have 1 W and 5 D's left in the box and

the probability that the fifth draw will be a D will be 5/10. Therefore, the probability of the sample WWWWD, in the order specified, is

$$\frac{5}{10} \times \frac{4}{9} \times \frac{3}{8} \times \frac{2}{7} \times \frac{5}{6} = \frac{600}{30{,}240} = \frac{5}{252}$$

If we shift the D to any other position in the sequence, we could show, in the same manner, that the probability of this sequence is the same as when the D is in the last position. It is clear, since D can take 5 different positions, and since the 5 sequences are mutually exclusive, that the probability of drawing a sample with 4 W's and 1 D is

$$\frac{(5)(5)}{252} = \frac{25}{252}$$

The probability of obtaining, in the order specified, WWWDD, is

$$\frac{5}{10} \times \frac{4}{9} \times \frac{3}{8} \times \frac{5}{7} \times \frac{4}{6} = \frac{1{,}200}{30{,}240} = \frac{10}{252}$$

and again this probability remains unchanged for all possible permutations of the 3 W's and 2 D's. The number of such permutations will be given by formula (2.4) and is equal to

$$\frac{5!}{3!\,2!} = 10$$

The probability, therefore, of obtaining a sample with 3 W's and 2 D's is

$$\frac{(10)(10)}{252} = \frac{100}{252}$$

Using the same methods, we find that the probability of obtaining a sample with 2 W's and 3 D's is 100/252; the probability of obtaining a sample with 1 W and 4 D's is 25/252; and the probability of obtaining a sample with 0 W's and 5 D's is 1/252.

THE BINOMIAL POPULATION

A population in which there are only two classes of observations is called a *binomial population.* We may arbitrarily assign the value of $X = 1$ to the observations in one of the two classes and the value of $X = 0$ to the observations in the other class. We let the number or frequency of observations in the population with the value of 1 be represented by F_1 and the number or frequency of observations with the value of 0 be represented by F_0. Then the total number of observations in the population will be equal to $N = F_1 + F_0$.

Table 3.1 shows a binomial population in which $F_1/N = .5$ and $F_0/N = .5$. We have intentionally not specified N, the number of observations in the population, since the points we wish to develop with respect to the binomial population are completely independent of N. In other words, we are not restricted in our discussion to a consideration of the

Table 3.1 Computation of the Mean and Variance of a Binomial Population

(1) Class	(2) $\frac{F}{N}$	(3) X	(4) $\frac{FX}{N}$	(5) $X - m$	(6) $(X - m)^2$	(7) $\frac{F(X - m)^2}{N}$
Wet	.5	1	.5	.5	.25	.125
Dry	.5	0	.0	−.5	.25	.125
$\sum$	1.0		.5	.0		.250

difference between finite and infinite populations. The distinction between finite and infinite populations is of importance only when we are concerned with the selection of a sample from a population. Our interest, for the moment, is in the characteristics of the binomial population, and the points to be made will be valid for both finite and infinite populations.

Mean of a Binomial Population

We define the mean of a population as

$$m = \frac{\sum X}{N} \tag{3.1}$$

or, if the observations have been arranged in the form of a frequency distribution, as

$$m = \frac{\sum FX}{N} \tag{3.2}$$

where m = a population mean
F = a population frequency
N = the number of observations in the population

For the population distribution of Table 3.1, we have $\sum FX/N = .5$, and this is the mean of the population. We also note that, for the binomial population,

$$m = \frac{\sum FX}{N} = \frac{F_1}{N} = P \tag{3.3}$$

or the proportion of the observations in the population in the class assigned the value of 1. The proportion of the observations in the population in the

class assigned the value of 0, may be designated by Q and it is obvious that $Q = 1 - P$.

Variance and Standard Deviation of a Binomial Population

We define the *variance of a population* as

$$\sigma^2 = \frac{\sum(X - m)^2}{N} \tag{3.4}$$

or, if the observations have been arranged in a frequency distribution, as

$$\sigma^2 = \frac{\sum F(X - m)^2}{N} \tag{3.5}$$

The square root of the variance is called the standard deviation. The *standard deviation of a population* will thus be given by

$$\sigma = \sqrt{\frac{\sum(X - m)^2}{N}} \tag{3.6}$$

or by

$$\sigma = \sqrt{\frac{\sum F(X - m)^2}{N}} \tag{3.7}$$

For the binomial population of Table 3.1, the variance is given by the sum of column (7) and is equal to .25. Thus, $\sigma = \sqrt{.25} = .5$.

It can also be shown that the standard deviation of the binomial population is equal to

$$\sigma = \sqrt{PQ} \tag{3.8}$$

where P and Q refer to the proportions in the population falling in the classes assigned the values of 1 and 0 respectively. For the population of Table 3.1, $P = .5$ and $Q = .5$. Thus, $\sigma = \sqrt{(.5)(.5)} = .5$, which is the same value we obtained by direct calculation in Table 3.1.

We emphasize again that the formulas given in this section are general and hold for *any* binomial population, finite or infinite, and for any value of P. We shall find formulas (3.3) and (3.8) of considerable value in subsequent discussions of methods useful in the evaluation of outcomes of research based upon sampling from a binomial population.

STATISTICS AND PARAMETERS

When we have a sample of n observations and obtain some measure based upon the sample, the resulting measure is called a *statistic*. For a given sample, the statistic of interest may be a mean, the frequency or

number of observations in a given class or with a given value, the proportion of observations in a given class or with a given value, a standard deviation, or any other measure that is based upon the sample set of n observations. When we have available a population of N observations, we may calculate similar measures, based upon the complete population. We distinguish these measures, based upon a complete population, from statistics based upon samples, by referring to the population measures as *parameters*. Thus, the mean of a sample of n observations would be a statistic, whereas the corresponding mean of the complete population would be a parameter. To make clear whether we are concerned with parameters or statistics, we shall use different symbols for them. Table 3.2 gives some

Table 3.2 Symbols Used When Referring to Measures Based Upon Samples and to Those Based Upon Populations

Measure	Sample	Population
Frequency	f	F
Number of observations	n	N
Proportion	p	P
Mean	$\bar{X}$	m
Variance	s^2	σ^2
Standard deviation	s	σ

of the different symbols we shall use in referring to measures obtained from samples and the corresponding measures based upon populations.

FREQUENCY AND SAMPLING DISTRIBUTIONS

When we classify observations in such a way as to show the number or frequency of observations in each class or with each value, we shall refer to this arrangement as a *frequency distribution*. A frequency distribution is simply a convenient way of showing how the observations are distributed among the various possible classes or values. We shall also be interested in the frequency distribution of a statistic based upon a number of different samples of n observations each. To distinguish the frequency distribution of a statistic from that of the individual observations, we shall refer to the frequency distribution of a statistic as a *sampling distribution*. In particular, the kind of sampling distribution we shall be interested in is a random sampling distribution. By a *random sampling distribution* we shall mean the sampling distribution of a statistic based upon random samples of n observations each drawn from some specified population. Random sampling involves the notion of random selection of observations as discussed previously.

EXAMPLE OF A FINITE POPULATION MODEL

In the case of the experiment with the farmer, what statistic is of interest? It should be one which will summarize the outcome of a particular sample of 5 selections. For the present, let us consider the proportion (p) of correct choices in each sample of $n = 5$. This proportion will be the number of times that a W can is selected, divided by the total number selected. Since, as we have already seen, a given sample may have 5, 4, 3, 2, 1, or 0 W's, and since $n = 5$, the possible values of p will be 1.0, .8, .6, .4, .2, and .0.

Table 3.3 summarizes the random sampling distribution of p for

Table 3.3 Calculation of Mean and Variance of p for Random Samples of $n = 5$ Drawn from a Finite Binomial Population with $N = 10$ and $P = .5$

(1) p	(2) $\frac{F}{N}$	(3) $\frac{Fp}{N}$	(4) $p - m$	(5) $(p - m)^2$	(6) $\frac{F(p - m)^2}{N}$
1 0	.00397	.00397	.5	.25	00099
.8	.09921	.07937	.3	.09	.00893
.6	.39683	.23810	.1	.01	.00397
4	.39683	15873	−.1	.01	.00397
.2	.09921	.01984	−.3	.09	.00893
.0	.00397	00000	− 5	.25	.00099
Σ	1.00002	50001	.0		.02778

samples of 5 drawn from a finite population of 10 in which the population proportion of W's is $m = P = .5$, and in which the population proportion of D's is $Q = 1 - P = .5$. The entries in column (2) of the table, headed F/N, are the theoretical relative frequencies or probabilities obtained earlier.[1]

Mean of the Distribution

The mean of the sampling distribution can be obtained by substituting the appropriate values from Table 3.3 in formula (3.2). In column (3) we give the products of column (2) and column (1). If we sum these products, we have the mean of the sampling distribution, and we see that this mean is equal to the population mean. Thus

$$m = P = \frac{\Sigma Fp}{N} \tag{3.9}$$

[1] As a result of rounding errors, columns (2) and (3) of Table 3.3 do not sum to 1.00 and .5, respectively, as they otherwise should.

where m is the population mean of the binomial distribution and, as shown earlier, is equal to P.

Variance and Standard Deviation of the Distribution

In column (4) of Table 3.3 we give the values of $p - m = p - P$, or the deviation of the sample values from the population mean. In column (5) the squares of these deviations are given. Multiplying each of these squared deviations by the corresponding values of F/N, given in column (2), we obtain the products shown in column (6). If we sum these products, we obtain the variance for the sampling distribution of p. Thus

$$\sigma_p^2 = \frac{\sum F(p - m)^2}{N}$$

or $\sigma_p^2 = .02778$, as given by the sum of the entries in column (6). Taking the square root of the variance, we have $\sigma_p = \sqrt{.02778} = .167$.

The standard deviation we have just obtained is the standard deviation of a sampling distribution, and any such standard deviation is called a *standard error*. A standard error refers to the variability of a statistic, as distinguished from a standard deviation which refers to the variability of individual observations.

THE GENERAL CASE FOR SAMPLES FROM FINITE POPULATIONS

Standard Error of p

If we have a binomial population with a given value for P, we have already seen that the population standard deviation will be

$$\sigma = \sqrt{PQ}$$

If random samples of size n are drawn from a finite binomial population without replacement and we find the value of p for each sample, these sample values of p will constitute a sampling distribution. The mean of this distribution will be equal to P, as we have already shown. The variance of the sampling distribution of p can be obtained directly if we know the population standard deviation, the number of observations in the population, and the sample size. Thus

$$\sigma_p^2 = \left(\frac{N - n}{N - 1}\right)\frac{\sigma^2}{n} \tag{3.10}$$

where N is the number of observations in the population and n is the sample size. Substituting in formula (3.10) with $N = 10$, $n = 5$, and $\sigma^2 = PQ = .25$, we have

$$\sigma_p^2 = \left(\frac{10 - 5}{10 - 1}\right)\left(\frac{.25}{5}\right) = \frac{.25}{9} = .02778$$

which is the same value we obtained by direct calculation in Table 3.3. The standard error of p will be given by the square root of formula (3.10). For the present problem we have $\sigma_p = \sqrt{.02778} = .167$, as before.

The important point is that formulas (3.9) and (3.10) are perfectly general. We can, for any finite binomial population with specified N and mean of P, determine, without making a single observation, that, if random samples of size n are drawn from this population, the mean of the sampling distribution of p will be equal to the population mean P. Furthermore, the standard error of p is known exactly and can readily be determined by means of formula (3.10).

Standard Error of f

It may be more convenient in some experiments to deal with the frequency (f) of correct discriminations rather than with the proportion (p) of correct discriminations. In the case of the experiment with the farmer, the frequency of correct choices is simply the number of W's in a given sample. Since we have assigned the value of 1 to each observation of W, and 0 to each observation of D, the frequency of W's is the sum of the values of the observations in the sample. It can be shown, in the general case, that the mean of the sampling distribution, when f is the statistic of interest, is equal to

$$m = nP \tag{3.11}$$

where n is the sample size and P is the population proportion in the class assigned the value of 1.

For example, we could multiply the entries in column (1) of Table 3.3 by n, so that instead of p, we would have the corresponding values of f. Then multiplying the values of f by the theoretical relative frequencies of column (2), and summing these products, we would find that the mean of the sampling distribution of f, in this problem, is $m = (5)(.5) = 2.5$.

It can also be shown, in the general case, that the variance of the sampling distribution of f, for samples drawn from a finite population, is given by

$$\sigma_f^2 = \frac{N - n}{N - 1} nPQ \tag{3.12}$$

For the data of Table 3.3, we have

$$\sigma_f^2 = \left(\frac{10 - 5}{10 - 1}\right)(5)(.5)(.5) = \frac{6.25}{9} = .69444$$

The standard error of f will be the square root of the value obtained by means of formula (3.12). Thus, for the present problem, we have $\sigma_f = \sqrt{.69444} = .833$.

EXAMPLE OF AN INFINITE POPULATION MODEL

In the second experiment described in the previous chapter, the population involved was also a binomial population, in that our observations could take only two values: the farmer could make a correct choice in a given pair by selecting the can that contained water or an incorrect choice by selecting the empty can. These observations corresponding to a given choice may, as before, be designated as W for a wet can or correct choice, and D for a dry can or an incorrect choice. The binomial population, in this instance, however, is regarded as infinite rather than finite. Consider, for example, a binomial population in which P, the population proportion of W's, is .5, and Q, the population proportion of D's, is also .5. If this population is finite and an observation is selected at random from the population and is not replaced, the values of P and Q for the observations remaining in the population will no longer be exactly equal to .5 and .5, respectively.

To simulate the population of observations in this experiment, we might consider a box in which 10 disks have the value of W and 10 have the value of D. Suppose we select at random one disk from this population. If the value of the observation is W, then the proportion of W's remaining in the box will be $\frac{4}{9}$ and the proportion of D's will be $\frac{5}{9}$, rather than .5 and .5, respectively. The probability of a D on the second draw, if the first observation is not replaced, will no longer be .5. In our analysis of the outcomes of the experiment, however, it is clear that the probability of obtaining a W on each trial was constant and equal to .5. That is because each trial consisted of a pair of cans, one W and one D, and the farmer's task was to choose one of the two.

We may simulate the infinite binomial population by replacing the disk in the box after each draw. In this way, the proportion of W's and D's in the population from which the sample is drawn will remain constant or unchanged. In the experiment with the farmer, the sample size was $n = 10$. In order to make the experiment more comparable to the first experiment, let us assume, however, that now the sample to be drawn from the infinite binomial population is 5 observations rather than 10.

By methods described previously, we obtain the sampling distribution of p, the proportion of W's, in random samples of $n = 5$ drawn from an infinite binomial population in which $P = .5$ and $Q = .5$. It should be clear that this binomial population will also have a standard deviation equal to $\sigma = \sqrt{PQ}$, even though it is infinite rather than finite.

The sampling distribution is shown in Table 3.4. We see that the mean of the sampling distribution of p, as given by the sum of column (3), is equal to the population mean or $m = P = .5$. The variance of p, as given

Table 3.4 Calculation of Mean and Variance of p for Random Samples of $n = 5$ Drawn from an Infinite Binomial Population with $P = .5$

(1) p	(2) $\frac{F}{N}$	(3) $\frac{Fp}{N}$	(4) $p - m$	(5) $(p - m)^2$	(6) $\frac{F(p - m)^2}{N}$
1.0	.03125	.03125	.5	.25	.0078125
.8	.15625	.12500	.3	.09	.0140625
.6	.31250	.18750	.1	.01	.0031250
.4	.31250	.12500	−.1	.01	.0031250
2	.15625	.03125	−.3	.09	.0140625
0	.03125	.00000	−.5	.25	.0078125
Σ	1.00000	.50000	.0		.0500000

by the sum of column (6), is now .050, rather than the value of approximately .028, which we obtained when we sampled from a finite population

THE GENERAL CASE FOR SAMPLES FROM INFINITE POPULATIONS

Standard Error of p

If we draw random samples of size n from an infinite binomial population, the mean of the sampling distribution will be equal to P, the mean of the binomial population. The variance of p can be obtained directly from the population standard deviation and the sample size. Thus the variance of p is given by

$$\sigma_p^2 = \frac{\sigma^2}{n} \tag{3.13}$$

and the standard error of p will be given by

$$\sigma_p = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{PQ}{n}} \tag{3.14}$$

For the data of Table 3.4, with $\sigma^2 = PQ$ and $n = 5$, we have, by means of formula (3.13), $\sigma_p^2 = .25/5 = .05$. The standard error of p will be the square root of the variance. Thus, $\sigma_p = \sqrt{.05} = .224$.

Standard Error of f

Instead of dealing with the sampling distribution of p, we could have analyzed the sampling distribution of f for the infinite case. We know that any given value of p is equal to f/n and, therefore, any given f is equal to np. The variance of the sampling distribution of p has been shown to be $\sigma_p^2 = \sigma^2/n = PQ/n$. If each value of p in the sampling distribution is

multiplied by n, and if the variance of this new distribution is obtained, it will be equal to n^2 times the variance of p. Thus

$$\sigma_f^2 = n^2\sigma_p^2 = n^2\frac{PQ}{n} = nPQ \tag{3.15}$$

and

$$\sigma_f = \sqrt{nPQ} \tag{3.16}$$

FINITE POPULATION CORRECTION FACTOR

Formulas (3.13) and (3.15), for the infinite case, differ from formulas (3.10) and (3.12), respectively, for the finite case, only by the factor, $(N - n)/(N - 1)$. This factor is a correction factor for sampling from a finite population. It is clear that if n, the sample size, is held constant, then, as N becomes indefinitely large, the correction factor approaches the limiting value of 1.00. In general, it can be said that if the ratio of n/N is less than $\frac{1}{5}$, then the standard error formula for the infinite population model will give a satisfactory approximation of that for the finite population model.

BINOMIAL EXPANSION

The sampling distribution of p for random samples of size n drawn from an infinite binomial population can be obtained directly by expanding the binomial $(P + Q)^n$, where P is the population proportion in one class, $Q = 1 - P$ is the proportion in the other class, and n is the sample size. The sampling distribution is obtained by substitution of the appropriate values of P, Q, and n in the following

$$(P + Q)^n = P^n + nP^{n-1}Q + \frac{n(n-1)}{1 \times 2}P^{n-2}Q^2 + \frac{n(n-1)(n-2)}{1 \times 2 \times 3}P^{n-3}Q^3 + \cdots + Q^n$$

Substituting in the above with $P = .5$, $Q = .5$, and $n = 5$, we have

$$(.5 + .5)^5 = (.5)^5 + 5(.5)^4(.5) + 10(.5)^3(.5)^2 + 10(.5)^2(.5)^3 + 5(.5)(.5)^4 + (.5)^5$$

and the successive terms give the theoretical relative frequencies of p equal to 1.0, .8, .6, .4, .2, and .0, respectively. These are exactly the same values as those given in column (2) of Table 3.4.

APPLICATIONS OF THE MODELS IN RESEARCH

The outcomes of many research problems can be evaluated in terms of random sampling from a binomial population, either finite or infinite. In general, the experimental design is one in which there are two classes of stimulus materials and the task set for the subject is the selection of those materials belonging to one of the two classes.

We might, for example, be interested in determining whether or not subjects can discriminate fresh orange juice from frozen orange juice, whether they can distinguish between two different cola beverages, two brands of cigarettes, or two brands of beer. In other cases, we may wish to determine whether subjects can distinguish between handwriting specimens of males and females or between photographs of individuals belonging to two different nationality groups. Still other research problems may involve determining whether subjects can distinguish between two different tones, or between two different intensities of sound.

In a psychophysical experiment, we may be interested in determining by how much one weight must differ from a standard weight before a subject can detect significantly better than by chance that the weight is heavier than the standard. Or we may wish to find out how salty a solution must be before it can be discriminated significantly better than by chance from a plain solution. Other applications will occur to the reader.

It should be emphasized that the methods described are perfectly general and do not require that P be .5, as in the two experiments with the farmer. We may have a population in which P is $\frac{1}{3}$, $\frac{1}{5}$, or any other value. The first experimental design, for example, could be modified by having 4 W cans and 6 D cans. The task set for the farmer would then be to select the set of 4 W cans from the complete set of 10. In this case, P would be .4, rather than .5. Similarly, in the second experiment, we might have arranged the cans in triplets, rather than in pairs, so that each triplet contained 1 W can and 2 D cans. The farmer would then be asked to select the W can in each set of 3. The analysis would proceed in the same manner, except that we would now have $P = \frac{1}{3}$ rather than $\frac{1}{2}$.

Knowing that we have an experiment which involves random sampling from a binomial population with known or assumed value of P, then, as we have seen, we can quickly and easily determine σ_f or σ_p for any specified value of n. In the next chapter we shall show how we can use σ_f or σ_p in evaluating the outcome of the experiment.

One other point should be mentioned. We have considered applications of the methods only to those experiments in which the stimulus materials can be divided into two classes. In the next chapter we shall see that, with only slight modifications, the methods of analysis can be extended to those experiments in which the stimulus materials can be divided into three or

more classes. In these experiments, the task set for the subject is to assign the materials to one of three or more classes.

QUESTIONS AND PROBLEMS

1. Write down the successive terms of $(\frac{1}{3} + \frac{2}{3})^6$.

2. We have a population of 8 disks, identified by the numbers 1, 2, 3, $\cdots$, 8. (*a*) If a random sample of 4 observations is drawn without replacement, what is the probability that the sample will include Disk 8? Show how you arrive at your answer. (*b*) How many possible different samples of 4 observations can be drawn from the population? (*c*) What is the probability of obtaining a sample which contains Disks 1, 8, 4, and 2? (*d*) If a sample of 5 is drawn without replacement, what is the probability of drawing Disks 8, 3, 1, 2, and 5 in that order? (*e*) What is the probability of drawing a sample containing Disks 8, 3, 1, 2, and 5, if the order is ignored?

3. We have a finite population of 8 disks. On 4 of the disks we have the letter A and on the other 4 the letter B. If a random sample of 4 observations is drawn, without replacement, show the probability of obtaining: (*a*) 4 A's; (*b*) 3 A's and 1 B; (*c*) 2 A's and 2 B's; (*d*) 1 A and 3 B's; (*e*) 4 B's.

4. Given an infinite binomial population with $P = \frac{1}{3}$ and $Q = 1 - P$. Random samples of $n = 15$ observations are drawn from the population. (*a*) What is the variance of the sampling distribution of p? (*b*) What is the variance of the sampling distribution of f?

5. Given a finite population of $N = 45$ observations with $P = \frac{1}{3}$ and $Q = 1 - P$. Random samples of $n = 15$ are drawn from this population without replacement. (*a*) What is the variance of the sampling distribution of p? (*b*) What is the variance of the sampling distribution of f?

6. Following the procedures outlined in Table 3.1, show that the population variance of the binomial population with $P = \frac{1}{3}$ and $Q = 1 - P$ will be equal to $\sigma^2 = (\frac{1}{3})(\frac{2}{3})$.

7. Following the procedures outlined in Table 3.3, show the sampling distribution of p for samples of $n = 4$ observations drawn from an infinite binomial population in which $P = \frac{1}{2}$ and $Q = 1 - P$.

8. Distinguish between (*a*) a frequency distribution and a sampling distribution; (*b*) a statistic and a parameter; (*c*) a standard deviation and a standard error; (*d*) a finite and an infinite population.

9. Under what conditions can we consider a sample of $n = 10$ observations drawn from some defined population to be a random selection from the population?

10. We have a discrimination experiment in which the probability of a correct discrimination is $\frac{1}{4}$. Subjects are to be given $n = 48$ independent trials. Assume the null hypothesis is true, that is, that a correct discrimination is a matter of chance. (*a*) What is the expected value or mean of the sampling distribution of f? (*b*) What is the standard error of f?

11. Define, briefly, each of the following terms:

binomial expansion	finite population correction factor
binomial population	variance

4

APPROXIMATION OF THE PROBABILITIES ASSOCIATED WITH SAMPLING FROM A BINOMIAL POPULATION

INTRODUCTION

Both of the experiments involving the farmer from Whidbey Island were limited, it was suggested, since, if the farmer could only do slightly better than chance, 5 observations would not be sufficient to result in a significant outcome. Thus, we might fail to reject the null hypothesis when it is in fact false, thereby making a Type II error. It was also suggested that, other things being equal, we could decrease the probability of a Type II error by increasing the number of observations made. But, as we increase the number of observations, we face the problem of evaluating the outcome of the experiment, and the methods of the previous chapter, which give us the exact distribution of the possible outcomes, become extremely laborious. It is fortunate that the probabilities associated with random sampling from a binomial population can be *approximated* quite satisfactorily by means of the table of the normal curve.

Since we have already obtained the exact probabilities for samples of 5 drawn from a finite and from an infinite population, we shall take these same two cases to illustrate the approximation method. With the exact probabilities available as a standard, we shall gain some notion of how well the approximation method works in these two specific cases.

THE UNIT NORMAL DISTRIBUTION

The normal distribution is one of the most useful distributions in statistical analysis. One reason why this is so is that the *random* sampling distribution of many statistics is approximately normal in form. In our discussion of the normal distribution, we shall assume that we are dealing

with the distribution of a continuous variable with known mean m and known standard deviation σ. The concept of a normal distribution applies to the shape or form of the distribution, however, and not to the specific mean and standard deviation of the distribution. Two or more distributions may both be normal in form and still differ with respect to their means and standard deviations. We can, however, shift the mean of any distribution from a specific value to $m = 0$. We do this by expressing each value of X as a deviation from m to obtain

$$x = X - m \tag{4.1}$$

and, since $\sum x = 0$, the mean of the distribution on this transformed scale will be equal to 0.

We can also transform any distribution to a new scale for which the unit of measurement is the standard deviation of the distribution. We do this by dividing each value of $x = X - m$ by the standard deviation of the distribution. Thus

$$z = \frac{X - m}{\sigma} \tag{4.2}$$

and it can readily be shown that any distribution transformed into z values by means of formula (4.2) will have, on the transformed scale, a mean of 0 and a standard deviation of 1. Furthermore, if X is normally distributed, then z will also be normally distributed. When this is the case, we shall refer to z as a *normal deviate*.

The various frequencies with which the values of a variable occur in a population of N observations can also be expressed as theoretical relative frequencies or proportions by taking $P = F/N$, and since

$$\sum P = \frac{\sum F}{N} = \frac{N}{N} = 1$$

we can make the area under the curve corresponding to a normal distribution equal to unity, regardless of the particular number of observations involved. We thus have an expression for the normal distribution that is independent of N, m, and σ. The normal distribution, in this form, is called the *unit normal distribution*, or *standard normal distribution*, and is applicable to any distribution that is normal in form, regardless of the particular mean, standard deviation, and N of the distribution. This theoretical distribution, since it is continuous, can be represented by a curve, and the equation for the unit normal curve is

$$y = \frac{1}{\sqrt{2\pi}} e^{-(\frac{1}{2})z^2} \tag{4.3}$$

where y = the height of the curve at any given point along the base line
π = 3.1416 (rounded), the ratio of the circumference of a circle to its diameter
e = 2.7183 (rounded), the base of the natural system of logarithms
$z = (X - m)/\sigma$

The area under this curve is equal to 1.00.

Table III in the Appendix is a table of the unit normal curve. Column (1), headed z, gives the distance on the base line from X to m in standard deviation units. The second column gives the proportion of the total area between the ordinates erected at m and z. The third column gives the area in the *larger* segment of the curve, and the fourth column gives the area in the *smaller* segment. The fifth column gives the value of y corresponding to the value of z, as obtained from formula (4.3) Since the normal curve is symmetrical, the tabled values are given only for positive values of z.

We illustrate the various relations described above in Figure 4.1 with $z = -1.65$. The proportions entered in the figure were obtained from Table III. Thus we see that .45 of the total area is contained between $z = -1.65$ and the mean which is equal to 0. We see that .95 of the total area falls above $z = -1.65$, that is, in the larger segment of the curve, and .05 of the total area falls below $z = -1.65$, that is, in the smaller segment.

It is of importance to observe that the proportions given in Figure 4.1, and obtained from Table III, correspond to various values of $\sum F/N$. The value of .05, for example, corresponding to the proportion of the total area falling below $z = -1.65$, means that of the total number of observations, .05 would fall to the left of $z = -1.65$. If we cumulate the frequencies

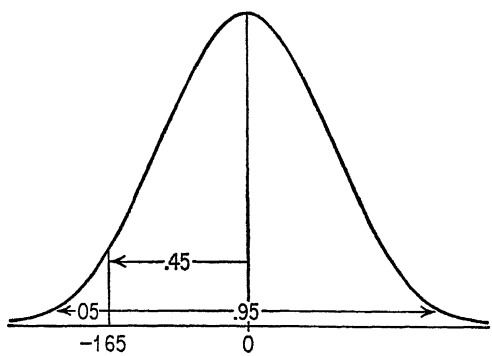


Figure 4.1 A normal distribution showing the proportion (.05) of the total area falling below $z = -1.65$, the proportion (.95) of the total area falling above $z = -1.65$, and the proportion (.45) of the total area falling between the mean and $z = -1.65$.

from left to right in the distribution until we have .05 of the total number, this value will be reached when we come to $z = -1.65$ on the scale of measurement.

TWO EXAMPLES OF APPROXIMATING BINOMIAL PROBABILITIES

Figure 4.2 shows the exact distribution of f, the frequency of correct choices, for random samples of $n = 5$ observations each drawn from a finite binomial population in which the probability of a correct choice is $P = .5$. The figure is based upon the data of Table 3.3. Figure 4.3 gives the corresponding distribution when the population is infinite, and is based upon the data of Table 3.4. We already know that the mean of both sampling distributions is $m = nP = 2.5$. It is also apparent that both distributions are symmetrical about the population mean. Furthermore, we know that $\sigma_f = .833$ for the distribution of Figure 4.2 and $\sigma_f = 1.118$ for the distribution of Figure 4.3.

It is also known that the binomial population from which the samples have been drawn is not a normally distributed population. But, as we indicated earlier, the random sampling distribution of many statistics tends toward normality, regardless of the shape or form of the population from which the samples are drawn. Let us see how well we can approximate the exact probabilities associated with certain values of f, by assuming that the distribution of f is normal. If our assumption that f is normally distributed is justified, then we may regard

$$z = \frac{f - m}{\sigma_f} \tag{4.4}$$

as a normal deviate.

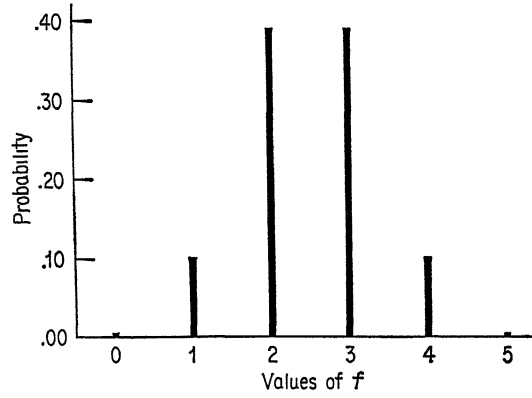


Figure 4.2 The sampling distribution of f for random samples of $n = 5$ drawn from a finite binomial population in which $N = 10$ and $P = .5$.

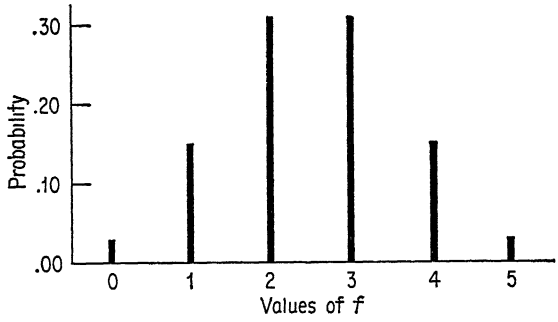


Figure 4.3 The sampling distribution of f for random samples of $n = 5$ drawn from an infinite binomial population in which $P = .5$.

Correction for Discontinuity[1]

One difficulty we face with formula (4.4) is that we know the distribution of f is discrete, whereas the normal distribution is concerned with variables that are continuous. We may correct for the discontinuity of f by treating the frequencies of correct selection in terms of an underlying parallel continuum. Thus, we shall regard a frequency of 5 correct choices as occupying an interval ranging from 4.5 up to 5.5; a frequency of 4 correct choices may be regarded as occupying an interval ranging from 3.5 up to 4.5, and so on.

Finite Population Example

Consider first the distribution of f for the finite population model. What is the probability of obtaining $f \geqq 5$ correct choices in a random sample of $n = 5$ from a normal distribution with $m = 2.5$ and $\sigma_f = .833$? Correcting f for discontinuity by taking its lower limit, 4.5, the probability will be given by the area in the right tail of the unit normal curve, that is, to the right of the normal deviate corresponding to $f = 4.5$. Assuming f to be normally distributed, we substitute in formula (4.4) and get

$$z = \frac{4.5 - 2.5}{.833} = 2.40$$

and from the table of the unit normal curve, we find the probability is .008. The exact probability, as obtained earlier, is .004.

Similarly, to obtain the probability corresponding to $f \geqq 4$, we find

$$z = \frac{3.5 - 2\,5}{.833} = 1.20$$

[1] This correction is also referred to as a "continuity" correction.

and the table of the unit normal curve gives this probability as .115. The exact probability of $f \geqq 4$, as obtained earlier, is .103.

Infinite Population Example

Consider now the exact probabilities for the infinite model and the corresponding probabilities as obtained by means of the normal curve approximation. We would now have for $f \geqq 5$

$$z = \frac{4.5 - 2.5}{1.118} = 1.79$$

and the probability, as obtained from the table of the unit normal curve is .037. The exact probability, as obtained earlier, is .031. For $f \geqq 4$, we have

$$z = \frac{3\,5 - 2.5}{1.118} = .89$$

and the probability as obtained from the table of the unit normal curve is .187. The exact probability, as obtained earlier, is also .187.

It is clear that in the cases examined the error introduced into the test of significance by the assumption of normality is not considerable. Indeed, the probabilities obtained by the normal curve test approximate quite well the exact probabilities.

Shape of the Sampling Distribution

It may be emphasized that one of the reasons why the normal curve approximations are as good as they are in the two cases described is that in random sampling from a binomial population with $P = Q$, the sampling distribution of f (and p) will be symmetrical. However, if P and Q are not equal, the sampling distribution of f (and p) will not be symmetrical but skewed. If P is larger than Q, the sampling distribution will have a tail to the left; if P is smaller than Q, the sampling distribution will have a tail to the right. However, as n, the sample size, increases, the sampling distribution becomes more symmetrical, even when P is not equal to Q. As a general rule, we can say that as long as nP and nQ are both $\geqq 5$, we shall not be seriously in error in using the normal curve approximations in tests of significance involving the binomial population.

TESTING NULL HYPOTHESES

It is worth emphasizing again that, from the point of view of the experimentalist and the research worker, a test of significance is a means for arriving at a decision concerning the null hypothesis tested. If we have chosen $\alpha = .05$ as our significance level, and if the normal curve test results in a probability that is considerably smaller or considerably larger than .05,

the decision we shall make with respect to the null hypothesis is clear-cut. For example, if in one of the experiments described we had tested $f \geqq 5$, the normal curve test would have resulted in a probability of .008 rather than the exact value of .004. We would make the same decision concerning the null hypothesis with the normal curve test that we would have made if we had used the exact test. Similarly, if we had tested $f \geqq 4$, both the exact and the normal curve test would result in the same decision, despite the fact that the probabilities obtained by the two tests differ somewhat. The important question is whether or not the two tests result in the same decision concerning the null hypothesis, rather than whether or not the probabilities obtained by the two tests are precisely the same.

Thus, the experimenter's interest is not primarily in the exact probability associated with the test of a given hypothesis, but rather in whether or not this hypothesis is to be regarded with suspicion. For this purpose, it does not seem likely that he will be led astray seriously if the normal curve test is applied instead of the exact test, as long as both nP and nQ are equal to or greater than 5, if the test is made by applying a correction for discontinuity and if the obtained probability is sufficiently small or sufficiently large to indicate that a conclusion concerning significance will not be changed if the test is made by means of exact methods. On the other hand, if the probability obtained by means of the normal curve test is of borderline significance, say between .07 and .03 when $\alpha = .05$, then the exact test may be applied and the decision to reject or not to reject the hypothesis may be made upon the basis of the probability obtained by the exact test.

TEST OF SIGNIFICANCE OF p

The normal curve test made by means of formula (4.4) is for the statistic f, the frequency of correct choices. In this instance we have $m = nP$ in the numerator and $\sigma_f = \sqrt{nPQ}$ in the denominator. If we desire to evaluate $p = f/n$ rather than f, then we may divide both numerator and denominator of formula (4.4) by n. Thus

$$z = \frac{\frac{f}{n} - \frac{nP}{n}}{\sqrt{\frac{nPQ}{n^2}}} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{p - P}{\sigma_p} \tag{4.5}$$

THE MATCHING PROBLEM

Suppose that we have an experiment in which we ask a subject to classify a set of stimulus materials into more than two categories. For

example, suppose that we have 45 test profiles on the Minnesota Multiphasic Personality Inventory (MMPI) obtained from patients at a mental hospital. Let us assume that 15 of these test records were obtained from male schizophrenics, 15 from male manic-depressives, and 15 from male psychopathic personalities, and that the patients are of the same age and of comparable educational level.

Subjects in our experiment consist of graduate students in the clinical psychology training program at a given university. All of the subjects have been trained in the administration and interpretation of MMPI records. Each subject is informed of the nature of the distribution of the 45 records, that is, that 15 belong in each of the three diagnostic categories, and his task is to arrange the 45 profiles into three sets of 15 each, with each set of 15 corresponding to one of the 3 categories.

The 45 observations obtained from any one subject may be arranged in the form of Table 4.1. Each row of this table shows how the 15 test

Table 4.1 The Correct and Incorrect Matches Made by a Subject in Classifying 45 MMPI Profiles into 3 Categories of 15 Each—The Correct Matches Are Shown in the Lower Left to Upper Right Diagonal

	Psychopaths	Manics	Schizophrenics	Total
Schizophrenics	2	3	10	15
Manics	6	5	4	15
Psychopaths	7	7	1	15
Total	15	15	15	45

records assigned by the subject to a given category are in fact distributed between the three categories. All of the off-diagonal entries will be errors or wrong matches and the total number of correct matches will be given by the sum of the diagonal cells $7 + 5 + 10 = 22$. We shall use f, the total number of correct matches, as the statistic of interest.

As in the case of the farmer, when we asked him to classify the set of 10 cans into two groups of 5 each, we shall assume that we have a finite population of N observations. If we have $c = 3$ categories and $n = 15$ profiles in each category, then $N = nc = (3)(15) = 45$. Let P be the probability that the first profile selected is placed in the correct category and $Q = 1 - P$. Then $P = n/N$ or $15/45 = 1/3$, for the problem at hand, and $Q = 2/3$.

The frequency of correct matches is to be based upon $N = 45$ observations, and it is the sampling distribution of f that we are interested in. Under the null hypothesis that the assignment of the 15 test records to

each category is a matter of chance, the expected or average number of correct matches, the mean of the sampling distribution of f, will be

$$m = NP = n \tag{4.6}$$

and the variance of this distribution will be given by

$$\sigma_f^2 = \frac{N}{N-1} NPQ \tag{4.7}$$

For the problem at hand, we have

$$m = 45\frac{1}{3} = 15$$

and

$$\sigma_f^2 = \frac{45}{45-1} 45\frac{1}{3} \times \frac{2}{3} = 10.2273$$

Then $\sigma_f = \sqrt{10.2273} = 3.2$. To evaluate the number of correct matches for the data of Table 4.1, we find, using a correction for discontinuity,

$$z = \frac{21.5 - 15}{3.2} = 2.03$$

and from the table of the normal curve we find that the probability of $f \geqq 22$ is .02. With $\alpha = .05$, the null hypothesis would be rejected and we would conclude that the subject has made more correct matches than can reasonably be attributed to chance.

Formulas (4.6) and (4.7) are applicable only to the case where we have the same number of stimuli in each category and when the subject is informed of this condition. Mosteller and Bush (1954) give formulas for more general cases. For example, we may not have the same number of stimuli in each category; or we may not restrict the subject as to the number of stimuli to be placed in each category.

SIGNIFICANCE OF THE DIFFERENCE BETWEEN TWO PROPORTIONS

In many experiments our interest is in the difference between two proportions or two frequencies. For example, a group of n subjects may be divided at random into two groups of n_1 and n_2 subjects. One group may then be given intermittent reinforcement in a conditioning experiment and the other group reinforcement on each trial. After comparable training periods, a critical trial is given each subject to determine whether the

conditioned response under investigation occurs or not. The experimenter is interested in knowing whether or not the frequencies or proportions of response in the two groups are significantly different.

As another example, in a learning experiment a group of n rats may be divided at random into two groups of n_1 and n_2 rats. One group is then given a series of training trials in a maze under one set of experimental conditions. The other group is also given a comparable training period, but under a different set of experimental conditions. On the basis of a particular learning theory, the rats in the first group should, when placed in a new maze with only two paths to the goal box, select one of the paths more frequently than the rats in the second group. The difference in response is in accordance with the learning theory, but the experimenter wishes to know whether the difference is statistically significant.

To illustrate the methods involved in evaluating the outcomes of experiments such as those described above, let us assume that we have randomly divided 80 rats into two groups with $n_1 = 50$ in one group and $n_2 = 30$ in the other.[2] Group 1 has been subjected to one set of experimental conditions and Group 2 to another set of experimental conditions. The appearance of a particular choice, response, or behavior pattern on a critical test trial which is in accordance with a theory may be called a success. The nonappearance of this choice, response, or behavior pattern may be called a failure. The frequency of successes and failures for the two groups in the critical test trial are given in Table 4.2.

Table 4.2 Frequency of Failures and Successes for Two Randomized Groups of Rats on a Critical Test Trial

	Failure	Success	Total
Group 1	8	42	50
Group 2	12	18	30
Total	20	60	80

If Group 1 is expected to show, on the basis of the theory, a greater proportion of successes, it is obvious that the outcome of the experiment is in the direction predicted by the theory, for the proportion of successes in Group 1 is $p_1 = 42/50 = .84$, whereas the proportion of successes in Group 2 is $p_2 = 18/30 = .60$. The problem is to determine whether these two proportions differ significantly.

[2] In general, in comparing a difference between two groups, we are better off with an equal number of subjects in each group. We have deliberately made the n's unequal in this example in order to illustrate a more general application of the methods of analysis.

Standard Error

Let us assume, as a null hypothesis to be tested, that the two sample sets of observations have been drawn at random from a common binomial population in which $P = 60/80 = .75$ and $Q = 1 - P = .25$. Then, if we have two random and independently selected samples of n_1 and n_2 observations from this binomial population, the standard error of the difference between the two sample proportions will be

$$\sigma_{p_1-p_2} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}$$

$$= \sqrt{\frac{PQ}{n_1} + \frac{PQ}{n_2}}$$

or, since the numerators are the same,

$$\sigma_{p_1-p_2} = \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \tag{4.8}$$

For the present example, we obtain

$$\sigma_{p_1-p_2} = \sqrt{(.75)(.25)\left(\frac{1}{50} + \frac{1}{30}\right)} = .10$$

Test of Significance

Now, assuming that the sampling distribution of the difference between p_1 and p_2 is approximately normal, we have

$$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sigma_{p_1-p_2}} \tag{4.9}$$

or a normal deviate. Then, since our null hypothesis specifies that the samples are from a common binomial with $P_1 = P_2$ so that $P_1 - P_2 = 0$, for the present example, we have

$$z = \frac{.84 - 60}{.10} = 2.4$$

By reference to the table of the normal curve, we find that .0082 of the total area will fall to the right of an ordinate at $z = 2.4$. Thus, if the null hypothesis is true, we would expect to obtain a difference as large as the one observed, *and in the direction observed* ($p_1 > p_2$), only 82 times in 10,000 as a result of random sampling.

One-Sided and Two-Sided Tests

If we have no prior hypothesis about the direction of the difference between p_1 and p_2, then our test of significance should take into account both the probability of a positive difference ($p_1 > p_2$) and the probability of a negative difference ($p_1 < p_2$). A directional test of a hypothesis is referred to as a *one-sided* or *one-tailed* test and a nondirectional test is referred to as a *two-sided* or *two-tailed* test. The nature of these tests will be discussed in greater detail later. For the present, it is sufficient to point out that the probability for the two-sided test will be two times the probability for the one-sided test. Thus, for the two-sided test we have $2(.0082) = .0164$ as the probability of a difference in either direction of the magnitude observed.

In the present experiment, assuming we have made a two-sided test with $\alpha = .05$, the null hypothesis would be rejected. We conclude that p_1 and p_2 differ significantly.

Correction for Discontinuity

In dealing with the one sample case, we introduced a correction for discontinuity. As will be recalled, the correction was such as to reduce the obtained value of z. In the test of significance of the difference between two proportions, we may also introduce a correction for discontinuity. The correction is made in such a way as to reduce the difference between the two sample proportions, that is, the frequency for the sample with the larger value of p will have .5 subtracted and the frequency for the sample with the smaller value of p will have .5 added. In the present example, making the correction for discontinuity, we would have

$$p_1 = \frac{42.0 - .5}{50} = .8300 \qquad \text{and} \qquad p_2 = \frac{18.0 + .5}{30} = .6167$$

The difference between p_1 and p_2 will now be $.8300 - .6167 = .2133$. Dividing this difference by the standard error of the difference, which we have previously found to be equal to .10, we have $z = .2133/.10 = 2.133$ for the value of z corrected for discontinuity. By reference to the table of the normal curve, we find this value to be significant, and the hypothesis of random sampling from a common binomial population would still be rejected.

The correction for discontinuity, in the example under consideration, resulted in no change with respect to our attitude toward the significance of the outcome of the experiment. In critical cases, however, the failure to apply the correction may result in the rejection of the hypothesis tested, whereas the corrected value of z may not be significant.

A Suggested Rule for Using the Normal Curve Test

In using the normal distribution to obtain an approximation of the probabilities associated with a single sample drawn from a binomial population, we had a rule that nP and nQ should both be equal to at least 5. A similar rule may be stated for the use of the table of the normal curve in evaluating the difference between two samples. The methods we have described will give fairly good approximations as long as for both samples we have n_1P, n_1Q, n_2P, and n_2Q all equal to or greater than 5. When this is not the case, the probabilities obtained from the normal curve test may be in error, and we may wish to consider a test of significance which yields exact probabilities.

EXACT TEST FOR THE DIFFERENCE BETWEEN TWO PROPORTIONS[3]

Table 4.3 gives a schematic representation for the cell frequencies in a 2×2 table. In the notation of this table, the probability of any observed set of frequencies will be given by

$$\frac{(a+b)!\,(a+c)!\,(b+d)!\,(c+d)!}{n!} \times \frac{1}{a!\,b!\,c!\,d!}$$

Assume, for example, we have the observed distribution of Table 4.4 and we wish to determine whether $p_1 = \frac{6}{8} = .6667$ and $p_2 = \frac{1}{7} = .1429$ differ significantly or whether the two samples can be assumed to be random

Table 4.3 Schematic Representation for the Frequencies in a 2×2 Table

	Failure	Success	Total
Group 1	a	b	$n_1 = a + b$
Group 2	c	d	$n_2 = c + d$
Total	$a + c$	$b + d$	$n = n_1 + n_2$

Table 4.4 Frequencies of Failure and Success for Two Randomized Groups

	Failure	Success	Total
Group 1	1	2	3
Group 2	6	1	7
Total	7	3	10

[3] In applying the exact test a table of logarithms of factorials is of value in simplifying the computations. Various tables are also available which make the evaluation of the difference between the proportions by means of the exact test very easy. See, for example, Finney (1948), Latscha (1953), and Mainland, Herrera, and Sutcliffe (1956).

samples from a common binomial population with $P = 3/10$. Then, substituting in the above expression, we have

$$\frac{3!\,7!\,3!\,7!}{10!} \times \frac{1}{1!\,2!\,6!\,1!} = .175$$

The desired probability, however, involves not only the set of cell frequencies recorded in Table 4.4, but all other possible sets which are more extreme with the marginal totals remaining the same. Thus, we also need the probability for the set:

$$\begin{matrix} 0 & 3 \\ 7 & 0 \end{matrix}$$

The probability of this set will be equal to

$$\frac{3!\,7!\,3!\,7!}{10!} \times \frac{1}{0!\,3!\,7!\,0!} = \frac{3!\,7!}{10!} = .008$$

Adding the two probabilities we have just calculated we have $.175 + .008 = .183$ for a one-sided test of significance and, with $\alpha = .05$, this is a non-significant value.

Although we do not need the probabilities for the other possible sets of cell frequencies, we calculate them to show that, if we take all possible sets, the sum of the probabilities will be equal to 1.00. Thus, we have the following additional possible sets of frequencies:

$$\begin{matrix} 2 & 1 \\ 5 & 2 \end{matrix} \quad \text{and} \quad \begin{matrix} 3 & 0 \\ 4 & 3 \end{matrix}$$

The probability for the first set is

$$\frac{3!\,7!\,3!\,7!}{10!} \times \frac{1}{2!\,1!\,5!\,2!} = .525$$

and the probability for the second set is

$$\frac{3!\,7!\,3!\,7!}{10!} \times \frac{1}{3!\,0!\,4!\,3!} = .292$$

The sum of the probabilities for all possible sets of cell frequencies is $.008 + .175 + .525 + .292 = 1.000$.

The data of Table 4.4 definitely do not meet the suggested standard that n_1P, n_1Q, n_2P, and n_2Q all be equal to or greater than 5. However, we may try the normal curve approximation to determine how well it works in this particular case. By substitution in formula (4.8), we have

$$\sigma_{p_1-p_2} = \sqrt{(.3)(.7)\left(\tfrac{1}{3} + \tfrac{1}{7}\right)} = .3162$$

Then making continuity corrections, we have

$$p_1 = \frac{2.0 - .5}{3} = .5000 \qquad \text{and} \qquad p_2 = \frac{1.0 + .5}{7} = .2143$$

By substitution in formula (4.9), we obtain

$$z = \frac{.5000 - .2143}{.3162} = .9035$$

By linear interpolation in the table of the normal curve between $z = .90$ and $z = .91$ we find that the probability of z equal to or greater than .9035 is approximately .183 and this estimate is the same as the probability we obtained by the exact test. In this instance, we could not ask for better correspondence between the approximate test based upon the normal curve and the exact test. This, of course, may not always be the case.

In practice, as we have stated before, the experimenter's interest may not be primarily in the exact probability of a test of significance, but rather in whether or not the hypothesis being tested is to be rejected or not at some specified level. For this purpose, the normal curve test will ordinarily be satisfactory. If the obtained probability is either very small or very large it is unlikely that the decision concerning the hypothesis will be changed by the application of the exact test. If the probability arrived at by the normal curve approximation is of borderline significance, and if any one of the values of n_1P, n_1Q, n_2P, or n_2Q is less than 5, then the exact test may be applied and the decision concerning the hypothesis can be made on the basis of the probability obtained by the exact test.

TEST FOR DIFFERENCE BETWEEN TWO PROPORTIONS WHEN THE SAME SUBJECTS ARE TESTED TWICE[4]

In some experiments we may have the same group of subjects tested twice. Suppose, for example, we have $n = 200$ subjects and they have been tested for the presence of a particular discriminatory response before (Test 1) and after (Test 2) experiencing a set of experimental conditions. We will thus have a pair of observations for each subject. If the discriminatory response is made, we shall call this a success (S), and if it is not, we shall call this a failure (F). Then there are two ways in which a subject may respond at the time of the first test (S_1 or F_1) and either of these two

[4] This test of significance has been described by McNemar (1947) as a test for the difference between two correlated proportions. For the case of three or more correlated proportions, see Cochran (1950).

ways may be followed by either of two ways at the time of the second test (S_2 or F_2). We thus have $2 \times 2 = 4$ possible patterns of response:

$$S_1F_2 \qquad S_1S_2 \qquad F_1F_2 \qquad F_1S_2$$

The Difference and Standard Error of the Difference Between p_1 and p_2

Let us assume that, in our particular experiment, the frequencies corresponding to the above patterns are 20, 80, 60, and 40, respectively. These frequencies are given in Table 4.5. In terms of the notation of Table 4.6, the proportion of successes on Test 1 will be

Table 4.5 Frequency of Failures and Successes for 200 Subjects Tested Before (Test 1) and After (Test 2) Experiencing an Experimental Condition

		Test 2		
		Failure	Success	Total
Test 1	Success	20	80	100
	Failure	60	40	100
	Total	80	120	200

Table 4.6 Schematic Representation of the Frequency of Failures and Successes When n Subjects Are Tested Twice

		Test 2		
		Failure	Success	Total
Test 1	Success	a	b	n_1
	Failure	c	d	$n - n_1$
	Total	$n - n_2$	n_2	n

$$p_1 = \frac{n_1}{n} = \frac{a+b}{n}$$

and the proportion of successes on Test 2 will be

$$p_2 = \frac{n_2}{n} = \frac{b+d}{n}$$

The difference between p_1 and p_2 will be

$$p_1 - p_2 = \frac{a+b}{n} - \frac{b+d}{n} = \frac{a-d}{n}$$

The test of significance, therefore, is concerned only with the distribution of the frequencies for patterns S_1F_2 and F_1S_2 or the frequencies in cells a and d of Table 4.6.

Now, if there is no difference between Test 1 and Test 2, then on the average we should expect a to be equal to d, that is, the frequency in cell a to be equal to the frequency in cell d. We can regard the set of $a + d$ observations as a sample from a binomial population in which the proportion of a's is equal to the proportion of d's. If we draw an observation at random from this binomial population, the probability of obtaining d will be $P = \frac{1}{2}$, and the probability of *not* d (a) will be $Q = 1 - P = \frac{1}{2}$.

Consider the sampling distribution of the frequency of d's. By formula (3.11) the mean of this distribution for a sample of $n = a + d$ observations will be

$$m = (a + d)P \tag{4.10}$$

The standard error of the distribution, by formula (3.16), for a sample of $n = a + d$ observations will be

$$\sigma_f = \sqrt{(a + d)PQ} \tag{4.11}$$

Test of Significance

If f is normally distributed, then

$$z = \frac{f - m}{\sigma_f}$$

is a normal deviate. Substituting (4.10) and (4.11) in the above formula, we have

$$z = \frac{d - (a + d)P}{\sqrt{(a + d)PQ}}$$

and since $P = Q = \frac{1}{2}$, then

$$z = \frac{d - (a + d)\frac{1}{2}}{\frac{1}{2}\sqrt{a + d}}$$

which may be simplified to

$$z = \frac{d - a}{\sqrt{a + d}} \tag{4.12}$$

With a correction for discontinuity, formula (4.12) becomes

$$z = \frac{|d - a| - 1}{\sqrt{d + a}} \tag{4.13}$$

Thus, the test of significance is easily made by means of formula (4.13).

For the data of Table 4.5, we have $d = 40$ and $a = 20$. Then

$$z = \frac{|40 - 20| - 1}{\sqrt{40 + 20}} = \frac{19}{7.746} = 2.45$$

From the table of the normal curve we find that the probability for the one-sided test is .0071 and for the two-sided test the probability is $2(.0071) = .0142$. Assuming we have made a two-sided test, with $\alpha = .05$, the null hypothesis is rejected. We conclude that the proportion of successes on Test 1 and the proportion on Test 2 differ significantly.

QUESTIONS AND PROBLEMS

1. In a taste discrimination experiment a subject is presented with two brands of frozen orange juice and fresh orange juice. His task is to select the fresh orange juice in each presentation of the three samples. He is given 15 trials and correctly selects the fresh orange juice in 9 of the 15 trials. Is the hypothesis that he is responding by chance tenable?

2. In a testing center it has been determined that the average test scorer is in error on 6 per cent of the papers scored. A new employee scores 500 papers during a given day, and it is found that 40 of his papers are scored incorrectly. We may regard the 500 papers which are scored as consisting of 500 trials of an event for which the probability of making an error in a single trial is .06. Has the new employee made a significantly large number of errors?

3. A subject is trained to push a key when the first of two tones which he hears is of greater intensity. The difference threshold for the subject is determined, and in a new series of trials two tones are sounded which differ in intensity, but for which the difference is below the threshold for the subject. The louder tone is randomly alternated with the weaker tone so that it sometimes appears first and sometimes second. The subject claims that he is unable to distinguish between the two tones—that he would have to guess. The experimenter tells him to go ahead and guess, and the subject does so for a series of 30 trials. If his judgment is a guess, then on each trial we may regard the probability of a correct guess as ½. What is the probability of 21 or more correct, if the null hypothesis is true?

4. A rat is trained to respond to the larger of two squares. The rat is now given a series of 40 trials with two circles differing in size. If we assume that the rat will respond to the two circles by chance, that is, that no preference will be shown for the larger of the two circles, then what is the probability of 26 or more responses to the larger of the two circles? Assume that the null hypothesis is rejected. Can this finding be interpreted as evidence of generalization from the previous training? Could it also be interpreted as evidence of learning in the present situation? Describe modifications of the experiment such that it would be possible to distinguish between the effects of learning in the present situation and the effects of generalization from the previous situation.

5. A child is presented with three boxes, of which two are of the same color and one is of a different color. Candy is placed under the box that is of the odd color and without the child's knowledge. He is then allowed to lift the boxes until he discovers the one which has the candy under it. The situation is now changed by using boxes of the same color but with two of the same size and one that is of a different size. Let us assume that there is no transfer of training from

his experience with the colored boxes to the boxes differing in size. The candy will always be placed under the box which differs in size from the other two and we shall assume that the probability of selecting this box is ⅓. What is the probability in a series of 18 trials that the child correctly selects the box with the candy under it 10 or more times? Suppose a significant result is obtained. Would it be just as logical to attribute this finding to learning in the present situation as to transfer from the previous situation? Describe modifications of the experiment such that it is possible to differentiate between learning in the present situation and transfer of training from the previous situation.

6. In a large midwestern university it is known that 62 per cent of the students are registered in the college of liberal arts. The campus daily draws a sample of 200 students for a public opinion poll and finds that in the sample there are 136 liberal arts students. If the sampling is random, how frequently would samples with 136 or more liberal arts students be expected by chance when the sample size is 200?

7. A child is seated at a table across from the experimenter. The child is shown a desired object, and this is placed at the end of the table at the child's right. Directly in front of the child across the table is a spot marked by an X. The child is blindfolded and asked to move a disk toward the spot. Assume that the probability of moving the disk to the right of the spot is equal to the probability of an error to the left. In a series of 20 trials, the child makes 14 right errors and 6 left errors. Does the child show a significant bias toward the position of the desired object? Would you wish to conclude, from the outcome alone, that the excess of right errors is the result of placing the desired object on the child's right? If so, then how would you answer the argument that the child might show a right-error bias if the desired object had been placed at the left? Describe modifications of this experiment such that the results can be interpreted as evidence of the desired object influencing the direction of the error.

8. Subjects are randomly assigned to two treatments A and B. After experiencing the treatments, both groups are given a critical test. A particular response is under investigation. In Group A, 24 out of 60 subjects make the response and in Group B, 8 out of 40 make the response. Can we conclude that these two proportions differ significantly?

9. Fifty-five subjects are given two shades of blue which differ only slightly with respect to saturation. They are asked to select the shade with the greater saturation and 36 of them make the correct selection. The same subjects are then retested with two shades of red which differ only slightly with respect to saturation. On this test it is found that 31 make the correct selection and that 24 of the 31 are also included in the 36 who made the correct selection in the test with blue. Can we conclude that the two proportions differ significantly?

10. The "Zeigarnik effect," which is concerned with the relative degree of recall of interrupted and completed tasks, has been studied by many psychologists. Tasks are presented to the subject and he is allowed to complete half of them and is interrupted on the other half. After the experimental session, the subject is asked to recall the tasks on which he has worked. A measure frequently used in such studies is the ratio (RI/RC) of the number of interrupted tasks recalled to the number of completed tasks recalled.

5

TESTS OF SIGNIFICANCE WITH THE χ^2 DISTRIBUTION

INTRODUCTION

The methods of analysis described in the previous chapter can be used in evaluating the outcomes of experiments in which we have one or two sets of observations from a binomial population. We now consider methods of analysis which can be used in evaluating the outcomes of experiments in which we have two or more classes of observations and/or in which we have more than two sets of observations.

For example, in a breeding experiment, a cross between two plants results in 352 seedlings. According to genetic theory, the seedlings should segregate into four types in the ratio of 9:3:3:1. The observed frequencies for the four types are 200, 72, 60, and 20, respectively. Are the data in accord with theory? Or do the frequencies deviate significantly from those expected on the basis of theory?

In a study of preferences, 120 subjects are presented with fresh, frozen, and canned orange juice. Each subject is asked to indicate the juice he prefers. The observed frequencies for the three juices are 60, 35, and 25, respectively. Can we conclude that these frequencies deviate significantly from a chance distribution?

In evaluating frequency distributions of the kind described, we shall make use of the χ^2 distribution.[1] For these problems, we may define χ^2 as

$$\chi^2 = \sum_{1}^{c} \frac{(f_i - F_i)^2}{F_i} \qquad (5.1)$$

where f_i is the observed frequency in the ith class, and F_i is a corresponding theoretical or expected frequency for that class, and the number of classes is equal to c. The theoretical frequencies are based upon a null hypothesis of interest. If the probability associated with the obtained value of χ^2 is small, then the null hypothesis will be rejected.

[1] For a further discussion of the χ^2 test, see Cochran (1954).

ONE SAMPLE WITH c CLASSES

The Preference Study

Consider the preference study with respect to fresh, frozen, and canned orange juice. The distribution of preferences for a sample of 120 subjects is given below:

	Fresh	Frozen	Canned
f	60	35	25

In this instance, the null hypothesis we may wish to test is that the three juices will be equally chosen. If this null hypothesis is true, the distribution of frequencies should be uniform, which is to say that $P_1 = P_2 = P_3 = \frac{1}{3}$ or that the probability of an observation falling in any one of the three classes is the same for all classes. Then, the expected frequency in each class for a sample of n observations will be

$$F = nP \tag{5.2}$$

or, for the problem at hand, where $n = 120$ and $P = \frac{1}{3}$

$$F = (120)(1/3) = 40$$

Then, by substitution in formula (5.1), we have

$$\chi^2 = \frac{(60 - 40)^2}{40} + \frac{(35 - 40)^2}{40} + \frac{(25 - 40)^2}{40} = 16.25$$

To find the probability of χ^2 equal to or greater than 16.25, when the null hypothesis is true, we make use of the table of χ^2, Table IV in the Appendix. To use Table IV, we must enter the table with the number of degrees of freedom (d.f.) associated with the obtained value of χ^2. The number of degrees of freedom may be regarded as the number of deviations $f_i - F_i$ that are free to vary. In the present problem, we note that $\sum_1^c f_i = \sum_1^c F_i$ so that $\sum_1^c (f_i - F_i) = 0$. Therefore, only $c - 1$ of the deviations are free to vary and this is the number of degrees of freedom associated with the χ^2 of 16.25. By reference to the table of χ^2, we find that for 2 d.f. a value of 9.21 or larger has a probability of .01. Thus, with $\alpha = .01$, the null hypothesis is rejected. We conclude that the distribution of preferences deviates significantly from a chance or uniform distribution.

The Genetic Experiment

For the genetic experiment, mentioned earlier, the observed frequencies for the four types of seedlings are as follows:

	Type 1	Type 2	Type 3	Type 4
f	200	72	60	20

According to theory, the seedlings should segregate in the ratio of 9:3:3:1. Thus, we have $P_1 = 9/16$, $P_2 = 3/16$, $P_3 = 3/16$, and $P_4 = 1/16$. We have $n = 352$ observations and the corresponding theoretical frequencies will be:

$$F_1 = (352)(9/16) = 198$$
$$F_2 = (352)(3/16) = 66$$
$$F_3 = (352)(3/16) = 66$$
$$F_4 = (352)(1/16) = 22$$

Then, with formula (5.1), we obtain

$$\chi^2 = \frac{(200 - 198)^2}{198} + \frac{(72 - 66)^2}{66} + \frac{(60 - 66)^2}{66} + \frac{(20 - 22)^2}{22} = 1.293$$

with $c - 1 = 3$ d.f. From the table of χ^2 we find that the obtained value of 1.293 is not significant. We conclude that the observed distribution does not deviate significantly from the distribution to be expected on the basis of genetic theory.

TWO OR MORE SAMPLES WITH c CLASSES

Rosenzweig's Study

Rosenzweig (1943) tested the recall of subjects for finished and unfinished tasks after they had worked on the tasks under differing sets of instructions. An "informal" group worked under the assumption that the experimenter was interested in studying work methods and that the ability of the subjects was not under investigation. A "formal" group worked on the same tasks under the impression that the problems were a kind of intelligence test. On some of the tasks both groups of subjects were interrupted, and on other tasks they were allowed to work until the problem was completed. At the end of the experiment, subjects in each group were asked to recall the tasks on which they had worked. We shall assume that the $n = 60$ subjects were divided at random into two groups of $n_1 = 30$ and $n_2 = 30$ subjects each. Table 5.1 gives the number of subjects in the

Table 5.1 Number of Subjects Showing a Tendency to Recall Finished and Unfinished Tasks and Number of Subjects Showing No Tendency of Differential Recall in Two Randomized Groups*

Group	Finished	Unfinished	No Tendency	Total
Informal	7	19	4	30
Formal	17	8	5	30
Total	24	27	9	60

* Rosenzweig (1943).

formal and informal groups who recalled a larger number of finished tasks, a larger number of unfinished tasks, or who showed no tendency in differential recall of the finished and unfinished tasks.

Test of Significance

If the difference in instructions to the two groups had no effect, we would expect the distribution of subjects in the classes of Table 5.1 to be similar for both groups. As a null hypothesis to be tested, we assume that both groups are from a common population in which the probabilities for each of the three classes of the table are $P_1 = 24/60$, $P_2 = 27/60$, and $P_3 = 9/60$, respectively. Then, since we have $n_1 = n_2 = 30$, the corresponding theoretical frequencies for each class for both groups will be:

$$F_1 = 30(24/60) = 12.0$$
$$F_2 = 30(27/60) = 13.5$$
$$F_3 = 30(9/60) = 4.5$$

If we let r = the number of rows or groups and c = the number of classes, as before, then, for problems of the kind described, we have

$$\chi^2 = \sum_1^r \sum_1^c \frac{(f_i - F_i)^2}{F_i} \tag{5.3}$$

where the double summation sign means that we must sum over each of the c classes for each of the r groups or rows.

In the cells of Table 5.2 we have entered the terms $(f_i - F_i)^2/F_i$

Table 5.2 The $(f_i - F_i)^2/F_i$ Terms for the Data of Table 5.1

Group	Finished	Unfinished	No Difference
Informal	$(7 - 12.0)^2/12.0$	$(19 - 13.5)^2/13.5$	$(4 - 4\,5)^2/4.5$
Formal	$(17 - 12\,0)^2/12.0$	$(8 - 13.5)^2/13.5$	$(5 - 4\,5)^2/4\,5$

corresponding to each of the f_i entries of Table 5.1. χ^2 is obtained by substitution in formula (5.3). Thus

$$\chi^2 = \frac{(7 - 12.0)^2}{12.0} + \frac{(19 - 13.5)^2}{13.5} + \cdots + \frac{(5 - 4.5)^2}{4.5} = 8.76$$

It may be observed in Table 5.2 that the deviations $(f_i - F_i)$ sum to zero in each row and each column of the table. Thus, only $(r - 1)(c - 1)$ of the deviations are free to vary. Accordingly, the χ^2 of formula (5.3) will have $(r - 1)(c - 1)$ degrees of freedom. For the present problem we have

a χ^2 of 8.76 with 2 d.f. From the table of χ^2 we find that, with $\alpha = .05$, our obtained value is significant. We therefore reject the null hypothesis and conclude that the two groups are not random samples from a common population with probabilities as given for the various classes. Examination of Table 5.1 shows that for the informal group there is a tendency for more unfinished tasks to be recalled, whereas for the formal group there is a tendency for more finished tasks to be recalled.

TWO OR MORE SAMPLES WITH $c = 2$ CLASSES

A Drug Study

In a mental hospital, a new drug at a standard dosage was tested. All male first admissions between the ages of 20 and 35 were given the drug. The observation recorded was whether or not the patient showed a reaction to the drug. Records were kept separately for each of nine months. The number of patients showing a reaction and the number showing no reaction are given in Table 5.3 for each of the nine months. The null hypothesis to

Table 5.3 Number of Nonreactors and Reactors to a Drug in Monthly Samples at a Mental Hospital

(1)	(2)	(3)	(4)	(5)	(6)
	Nonreactors	Reactors			
Months	f_1	f_2	n_i	f_1^2	f_1^2/n_i
January	18	32	50	324	6.480
February	20	25	45	400	8.889
March	22	20	42	484	11.524
April	19	19	38	361	9.500
May	14	22	36	196	5.444
June	21	19	40	441	11.025
July	22	21	43	484	11.256
August	16	20	36	256	7.111
September	10	20	30	100	3.333
$\sum$	162	198	360		74 562

be tested is that the groups tested each month are from a common population in which the probability of a reaction to the drug is $P = 198/360 = .55$ and the probability of a nonreaction is $Q = 1 - P = .45$.

Test of Significance

When we have r samples and only $c = 2$ classes of observations there is a simplified method for calculating χ^2. We take the column of frequencies

in Table 5.3 with the smaller total.[2] In the present instance, this is column (2), headed f_1, which shows the frequency of nonreactors in each group. We now square each of the f_1 values to obtain the entries in column (5). In column (6) we have divided each ${f_1}^2$ by n_i, the number of observations in the group. We then find the sums of the columns as shown at the bottom of the table and substitute in the following formula to obtain

$$\chi^2 = \frac{n^2}{\sum f_1 \sum f_2}\left[\sum \frac{{f_1}^2}{n_i} - \frac{(\sum f_1)^2}{n}\right] \tag{5.4}$$

where n = the total number of observations in all samples
$\sum f_1$ = the total number of observations in one of the two classes
$\sum f_2$ = the total number of observations in the other class
n_i = the number of observations in the ith group

Making the substitutions from Table 5.3 in formula (5.4), we obtain

$$\chi^2 = \frac{(360)^2}{(162)(198)}\left[74.562 - \frac{(162)^2}{360}\right] = 6.72$$

with degrees of freedom equal to $(r - 1)(c - 1) = (9 - 1)(2 - 1) = 8$. By reference to the table of χ^2, we find that the obtained value of 6.72 is not significant. We conclude that the various groups are from a common population in which the probability of a reaction to the drug is .55 and the probability of nonreaction is .45.

χ^2 As a Test of Independence

It is of some interest to consider the nature of the test of significance in this example. Randomization was not involved in the assignment of subjects to the various groups (months) and the treatment was the same for each group. If a significant value of χ^2 had been obtained, what would this mean? Statistically, it would mean that the proportions of reactors and nonreactors differed in the monthly samples. This, in turn, may indicate that the patients entering the hospital during certain months differed in some systematic way from the patients entering during other months or that the nature of the drug or its administration differed systematically between the months. A significant χ^2, in this instance, would tell us nothing about the conditions resulting in the differences in the proportions between the various months. The test of significance, in the absence of randomization, may be interpreted as providing an indication of whether or not

[2] It is not necessary to take the column of frequencies with the smaller total, but this simplifies the computations somewhat. If we choose to use the f_2 column rather than the f_1 column, we interchange the terms for f_1 and f_2 in formula (5.4).

the row variable (months) and the column variable (reaction or no reaction) are independent or associated. A nonsignificant value of χ^2 indicates that the two classifications are independent, whereas a significant value indicates that they are associated.[3] It is well known that if two variables are associated, this alone does not tell us which may be cause and which may be effect. We shall have more to say upon this point in the next section and in subsequent discussions.

TWO SAMPLES WITH $c = 2$ CLASSES

The Maier Study

A reasoning problem which involved clamping together two sticks so that the length was just sufficient to wedge the joined sticks between the floor and ceiling of an experimental room was used in an investigation by Maier (1945) The subjects were instructed to construct a hat rack from the materials supplied, and the solution was as described above, the projection of the clamp from the two sticks providing the necessary hook for

Table 5.4 Number of Men and Women Reaching No Solution or a Solution in a Reasoning Problem*

	No Solution	Solution	Total
Men	13	26	39
Women	26	10	36
Total	39	36	75

* Maier (1945).

hanging up a coat or hat. Men and women were used as subjects and they were tested under three different experimental conditions—the conditions involving different clues to the solution of the problem. The data given in Table 5.4 are the totals for all three conditions.

Test of Significance

Assume that the null hypothesis of interest is that the group of men and the group of women are from a common binomial population in which the probability of a solution is $P = 36/75 = .48$ and the probability of no solution is $Q = 1 - P = .52$. For the $r \times c = 2 \times 2$ table, that is, with

[3] For a discussion of various measures of association for cross classifications of the kind described in this chapter, see Goodman and Kruskal (1954, 1959).

two groups or rows and two columns or classes, we have the schematic representation shown in Table 5.5. Then, in the notation of this table

Table 5.5 Schematic Representation of Frequencies for $r = 2$ Groups and $c = 2$ Classes

	Failure	Success	Total
Group 1	a	b	$n_1 = a + b$
Group 2	c	d	$n_2 = c + d$
Total	$a + c$	$b + d$	$n = n_1 + n_2$

$$\chi^2 = \frac{n\left(|bc - ad| - \frac{n}{2}\right)^2}{(a + b)(c + d)(a + c)(b + d)} \tag{5.5}$$

where the factor $n/2$ is a correction for discontinuity.

Substituting the data of Table 5.4 in formula (5.5), we have

$$\chi^2 = \frac{75\left(|676 - 130| - \frac{75}{2}\right)^2}{(39)(36)(39)(36)} = 9.84$$

with 1 d.f.[4] According to the table of χ^2 our obtained value is significant.

Problems in the Interpretation of the Results

What may we conclude from this study? The treatment is the same for both groups. It is unlikely that the 39 men involved in the study constitute a random sample from any larger population of men and it is also unlikely that the 36 women constitute a random sample from any larger population of women. For the particular samples involved, there is evidence that one classification, sex, and the other classification, solution or no solution, are not independent. There is an association, in other words, between the row and column variables for this particular sample.

It is of importance to understand that the finding of an association between the two variables cannot be interpreted in the same manner as when we have an experiment in which different treatments are involved and the treatments are randomly assigned to subjects. With randomization and with a significant difference between treatment groups, we have a basis

[4] If χ^2 has 1 d.f., then $\chi^2 = z^2$ and it is possible to use the more complete table of the normal curve to find the probability associated with χ^2. For example, if $\chi^2 = 4.0$, with 1 d.f., then $z = 2.0$ or -2.0. From the table of the normal curve we find when $z = 2.0$, then $P = .0228$ and the probability associated with χ^2 will be $(2)(.0228) = .0456$.

for concluding that the observed difference between the two groups is the result of the difference in the treatments. With randomization, we expect individual differences (organismic variables) to be randomized over the treatments. In the present example, we would not wish to attribute the difference in the frequency of solutions and no solutions between the two groups to the one obvious way in which the two groups differ, that is, sex, since the groups may also differ in many other respects.

Perhaps the point we wish to make can be emphasized by considering some fictitious but possible conditions that may be present in the study under consideration. Suppose, for example, that every subject classified as a male was also blue-eyed and that every subject classified as a female was brown-eyed. Then it would also be true that we would have an association between eye color and failure and success in the problem. Or, we may not be wrong in assuming that each male subject wore loafers and each female subject wore brown and white saddle shoes. If this were the case, then we would also have an association between type of shoe worn and failure and success on the problem.

Some individuals might argue that we have no reason for believing that either eye color or type of shoe would be associated with failure and success. But it may also be argued that we have no reason for believing that sex would be associated with the outcomes of the study either. Under any circumstance, it is not likely that the sex classification is the only systematic way in which the two groups differ. In the absence of evidence to the contrary, any such systematic difference between the two groups would be as plausible (or implausible) an explanation of the results as the sex classification.

TEST OF TECHNIQUE

Nature of the Experiment

In an experiment concerning the influence of a particular drug upon a physiological response, the drug was to be tested at two levels of concentration. The drug was to be administered by injection and the experimenter was not sure of his technique, that is, if comparable groups were tested a second time whether or not he would obtain the same or comparable results. The experimental design provided for a test of the technique by repeating the complete experiment four times.

Subjects were divided at random into eight groups of 20 subjects each. Four of the groups were assigned at random to each level of the drug. The observation for each subject was the presence or absence of a specified reaction to the drug. The distributions of reactors and nonreactors in each group for each level of the drug are given in Table 5.6.

Table 5.6 Number of Nonreactors and Reactors in Randomized Groups with Two Levels of a Drug

	(1) Groups	(2) Nonreactors f_1	(3) Reactors f_2	(4) Total n_i	(5) f_1^2	(6) f_1^2/n_i
First level	1	10	10	20	100	5.00
	2	12	8	20	144	7.20
	3	8	12	20	64	3.20
	4	15	5	20	225	11.25
	Total	45	35	80		26.65
Second level	1	6	14	20	36	1.80
	2	8	12	20	64	3.20
	3	5	15	20	25	1.25
	4	6	14	20	36	1.80
	Total	25	55	80		8 05

Test of Experimental Procedure

Consider first only the four groups of subjects tested at the first level. If the experimenter's technique is under control, then we would expect to find the distribution of reactors and nonreactors in each group to be comparable from group to group. The reason for this is that randomization was involved in the assignment of subjects to the groups and each group received the same treatment. Using formula (5.4), we have

$$\chi^2 = \frac{(80)^2}{(45)(35)}\left[26.65 - \frac{(45)^2}{80}\right] = 5.43$$

and this is a nonsignificant value for 3 d.f. The null hypothesis that these four groups are from a common population in which the probability of a reaction to the drug at the first level is $P = 35/80 = .4375$ is tenable. If a significant value of χ^2 had been obtained in this instance, it would indicate that something was apparently wrong with the experimental technique. Possible explanations for this might be found in systematic differences in the manner of injection, in the dosage injected, or in some other aspect of the experimental technique. As it is, the nonsignificant value of χ^2 indicates that the proportions of reactors in the various groups are comparable.

Similarly, for the groups tested with the second level, we have

$$\chi^2 = \frac{(80)^2}{(25)(55)}\left[8.05 - \frac{(25)^2}{80}\right] = 1.11$$

and this is also a nonsignificant value for 3 d.f.

Test for Difference Between Levels

Since the tests of technique indicate that the groups tested at the first level are homogeneous and that the groups tested at the second level are also homogeneous, we may pool the results for each level to obtain Table 5.7. We now wish to test the null hypothesis that the groups tested at the

Table 5.7 The Pooled Results for the Data of Table 5.6

	Reactors	Nonreactors	Total
First level	45	35	80
Second level	25	55	80
Total	70	90	160

two different levels are from a common population in which the probability of a reaction is $P = 90/160 = .5625$. Using formula (5.5), we have

$$\chi^2 = \frac{160\left(|875 - 2{,}475| - \frac{160}{2}\right)^2}{(80)(80)(70)(90)} = 9.17$$

with 1 d.f. From the table of χ^2 we find that this is a significant value and we reject the null hypothesis. Examination of the data of Table 5.7 shows that a higher proportion of reactors are found at the second level than at the first level. Since randomization was involved in assigning subjects to the two levels of the drug, we have a basis for concluding that the observed difference between the two proportions is the result of the difference in the treatments, that is, the level of the dosage.

χ^2 WITH MORE THAN 30 d.f.

The table of χ^2 provides entries for degrees of freedom equal to 30 or less. For a larger number of degrees of freedom, we may find

$$z = \sqrt{2\chi^2} - \sqrt{(2)(\text{d.f.}) - 1} \tag{5.6}$$

The value of z obtained in formula (5.6) is approximately normally distributed with zero mean and unit standard deviation and thus may be considered a normal deviate to be evaluated by means of the table of the normal curve.

Suppose, for example, that we obtain a value of χ^2 from a table with $r = 10$ groups or rows and $c = 5$ columns or classes. We thus have $(10 - 1)(5 - 1) = 36$ d.f. If the obtained value of χ^2 is equal to 54.5, then

$$z = \sqrt{(2)(54.5)} - \sqrt{(2)(36) - 1} = 2.01$$

From the table of the normal curve, we find that the area in the right tail cut off at an ordinate erected at $z = 2.01$ is .0222. This is the probability for a directional or one-sided test. For the nondirectional χ^2 test, the probability will thus be (2)(.0222) = .0444. With $\alpha = .05$, the obtained $\chi^2 = 54.5$ would be regarded as significant.

QUESTIONS AND PROBLEMS

1. Kuenne (1946) studied transposition behavior in two groups of children who differed with respect to age. Group 1 consisted of 18 children ranging in age from approximately 34 to 46 months. Group 2 consisted of 26 children ranging in age from approximately 60 to 63 months. In the critical test trials, 3 of the children in Group 1 showed transposition behavior and 15 did not. In Group 2, the number showing transposition behavior in the critical test trials was 20, while 6 failed to meet the criterion. Can we conclude that the two proportions differ significantly?

Again, in this experiment, we must note that randomization was not involved in assigning subjects to the two groups. Age is an organismic variable and cannot therefore be randomly assigned to a subject. What bearing does this have upon the interpretation of the outcome of the study? Would you attribute the results to the age difference? If so, how would you answer the argument that the subjects may also differ with respect to important organismic variables other than age?

2. In the above problem, use the exact methods for the 2×2 table, described in the previous chapter, to determine the probability of the outcome and all other outcomes more extreme and in the same direction.

3. In a study by Hellman (1914) it is reported that of 20 breast-fed youngsters, 4 had normal teeth and 16 showed malocclusion. Of 22 bottle-fed youngsters, 1 had normal teeth and the other 21 showed malocclusion. Can we conclude that the two proportions differ significantly?

Since randomization is not involved in this study, what bearing would this have upon the interpretation of the result of the experiment if it had been significant? Is it possible that mothers who breast-feed their youngsters may also differ in other respects from mothers who bottle-feed their youngsters? What other variables might be associated with breast-feeding and bottle-feeding which, in turn, might be associated with malocclusion and normal teeth?

4. Records were kept of the number of students who left a university auditorium through each of three main exits. For a sample of 795 students the counts were as follows: Exit 1, 245 students; Exit 2, 200 students; Exit 3, 350 students. Can we conclude that the exits are equally popular?

5. Kendall and Smith (1939) have described the tests they applied to their tables of random numbers. All the numbers in the published tables were run off by one operator using an electrical device constructed for the purpose. One of the tests applied to the numbers drawn was the frequency test which consisted of

counting the frequencies of the digits from 0 to 9. Various sets of numbers were rejected, including this one:

Digit	f	Digit	f
0	1,083	5	1,007
1	865	6	1,081
2	1,053	7	997
3	884	8	1,025
4	1,057	9	948

Assuming randomness, the expected frequency for each digit is 1,000. Can we conclude that the probability for each digit is the same?

6. Hartman (1939) tested men and women with various solutions of phenylthiocarbamide. The solutions were numbered in terms of strength from 0 to 10, and the threshold was recorded as the concentration below which they first tasted the presence of phenylthiocarbamide. Since some subjects tasted the weakest solution 0, the threshold for these subjects was recorded as below 0, giving rise to 12 classes. The frequency distributions of the thresholds for 290 men and 314 women were as below:

	Frequency		
Strength	Men	Women	Total
10	15	42	57
9	35	52	87
8	46	38	84
7	31	30	61
6	23	19	42
5	13	17	30
4	9	6	15
3	7	5	12
2	10	10	20
1	13	19	32
0	25	33	58
Below 0	63	43	106

Can we conclude that threshold and sex classification are independent?

7. Records were kept at a university medical clinic of students who had attacks of influenza. Some of these students had been given vaccinations against influenza and others had not. The students were also classified in terms of whether they had a severe attack or a minor attack. The data are as follows:

	Minor Attack	Severe Attack
Vaccinated	98	40
Not vaccinated	30	82

Can we conclude that these two variables are independent?

In the absence of randomization, would we want to attribute the severity of the attack to the presence or absence of vaccination? What are some of the

possible systematic organismic differences that may exist between subjects who were vaccinated and those who were not?

Assume that a design could be worked out in which subjects would be randomly assigned to the vaccination and nonvaccination groups. What additional controls would be needed in this study? Would it make any difference if the physician who did the vaccinating also did the rating of the severity of the attack? How could this possible source of bias be controlled? A subject's knowledge of the fact that he has or has not been vaccinated might be of some importance. How could this be controlled? Should consideration be given to those subjects, vaccinated and nonvaccinated, who have no attacks?

8. Merritt and Fowler (1948) report a study in which the procedure was as follows. ". . . stamped, self-addressed, and sealed letters of two types were 'lost' by depositing them prominently but discreetly on sidewalks of various cities in the East and Midwest. Type A contained only a trivial message, while Type B contained, besides a message, a lead slug of the dimensions of a fifty-cent piece. The accompanying message indicated that the lead disk, as such, was of value to the addressee. Care was taken to drop the letters in locations sufficiently removed from one another to preclude the possibility of any one person finding more than one of the letters. All were put down in clear weather so that the envelopes would not become soiled and hence lose their appearance of value. Tests were made by night and day in both business and residential districts" (pp. 90–91).

Thirty-three letters of Type A were dropped and of these 28 were returned by the person picking them up. Of Type B, 158 letters were dropped and 86 of these were returned. Can we conclude that the probability of a Type A letter being returned is the same as the probability of a Type B letter being returned?

9. Define, briefly, each of the following terms:

expected frequency
test of independence
test of technique or experimental procedure

6

SIGNIFICANCE TESTS FOR THE CORRELATION COEFFICIENT

INTRODUCTION

One of the most frequently used statistics in psychological research is the product moment coefficient of correlation. The coefficient of correlation is a measure of the degree of linear association between two variables. The coefficient may be positive or negative in sign and ranges in value from -1.00 to 1.00.

A correlation coefficient may be computed whenever observations are paired. For example, if subjects are given two psychological tests, then each subject will have two scores, one on each test. The two scores for each subject constitute a pair of observations.

A positive correlation between two tests will be obtained when subjects who are above the mean on one of the tests also tend to be above the mean on the other test, whereas subjects who are below the mean on one of the tests also tend to be below the mean on the other test. A negative correlation, on the other hand, will be obtained when subjects who are below the mean on one test tend to be above the mean on the other test, whereas subjects who are above the mean on the first test tend to be below the mean on the second test.

Let one of the two variables for which the correlation coefficient is to be computed be symbolized by X and the other variable by Y. Then, using r to designate the correlation coefficient, we have

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} \tag{6.1}$$

where $\sum xy = \sum (X - \bar{X})(Y - \bar{Y})$, $\sum x^2 = \sum (X - \bar{X})^2$, $\sum y^2 = \sum (Y - \bar{Y})^2$, and the summation is over the n paired X and Y values. It will be convenient to refer to each pair of X and Y values as an observation.

SAMPLING DISTRIBUTION OF r

If we have a sample of n observations from a defined population, we can compute the correlation coefficient for this sample by means of formula

(6.1). If we draw an indefinitely large number of successive random samples of n observations each from the defined population and the value of r is computed for each sample, the distribution of these values will be the sampling distribution of r for samples of size n. If the sample values of r were normally distributed about the population value, and if the standard error of the distribution were known, we could proceed to test various hypotheses in the manner already familiar by means of the table of the normal curve.

However, if n is small, the sampling distribution of r will be decidedly skewed if the population correlation is moderately large. The degree of skewness will depend upon both n and the magnitude of the population correlation. The smaller the n and the larger the magnitude of the population correlation (either positive or negative), the greater the degree of skewness in the sampling distribution. For any given population value, as n increases, the skewness tends to decrease and the sampling distribution tends to become more symmetrical. For any given n, as the population value approaches zero, the sampling distribution of r also tends to become more symmetrical, but not necessarily normal.

THE t TEST OF THE HYPOTHESIS OF ZERO CORRELATION

To test the null hypothesis that the population correlation is zero, we make use of the table of t, Table V in the Appendix. As in the case of the table of χ^2, to use the table of t we must enter the table with the number of degrees of freedom available. If the null hypothesis that the population correlation is zero is true, then

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{6.2}$$

has a t distribution with $n - 2$ d.f. and can be evaluated by means of the table of t.

Suppose that a sample of $n = 11$ observations results in an r of .60. Then substituting in formula (6.2) we have

$$t = \frac{.60\sqrt{11-2}}{\sqrt{1-(.60)^2}} = 2.25$$

with $11 - 2 = 9$ d.f. By reference to the table of t, we find that for 9 d.f., a value of 2.262 has a probability of .025. This is the probability of obtaining a positive value of t equal to or greater than 2.262 and it corresponds to the probability of obtaining a positive value of r equal to or greater than .60, when the null hypothesis is true. The t distribution, like the z distribution, is symmetrical. Therefore, the probability of obtaining a nega-

tive value of t equal to or less than -2.262 is also .025, and this corresponds to the probability of obtaining a negative r of $-.60$ or less, when the null hypothesis is true. Thus, if we wish to make a nondirectional or two-sided test of significance, with $\alpha = .05$, the probabilities given by the column headings of Table V should be doubled. For the nondirectional test, and with $\alpha = .05$, we should be prepared to reject the null hypothesis if we obtain an absolute value of t equal to or greater than 2.262.

TABLE OF SIGNIFICANT VALUES OF r

By substituting the values of t from Table V and various values of n in formula (6.2), it is possible to solve for the values of r that would be significant at specified levels of significance. Table VI in the Appendix gives these values and by reference to this table one can quickly determine whether a given value of r based upon a given number of degrees of freedom $n - 2$ is sufficiently large to cause us to reject the hypothesis that the population correlation is zero. Table VI gives values of r that would be regarded as significant at the levels given by the column headings, if one-sided or one-tailed tests of significance are made. If a two-sided or nondirectional test is made, then the probabilities given by the column headings should be doubled.

While Table VI is convenient for testing the null hypothesis that a population correlation is zero, it is of no value in testing other hypotheses concerning the population value. Nor can formula (6.2) be used for this purpose. Suppose, for example, we have obtained $r = .45$ with $n = 42$ observations. With $\alpha = .05$, and making a two-sided test, we find that the tabled value of r which would be regarded as significant is .304. Thus, we would reject the null hypothesis of zero correlation. But suppose we wish to test, not the null hypothesis of zero correlation in the population, but rather the null hypothesis that the population correlation is some other specified value, say, .20. To test this null hypothesis, we make use of the z' transformation for r.

THE z' TRANSFORMATION FOR r

The value of z' for any given value of r is

$$z' = \frac{1}{2}[\log_e (1 + r) - \log_e (1 - r)] \tag{6.3}$$

where r is the observed value of the correlation coefficient. In order to make the z' values directly available without resort to a table of natural logarithms, values of r were substituted in formula (6.3) and the corresponding values of z' were found. These values are given in Table VII in

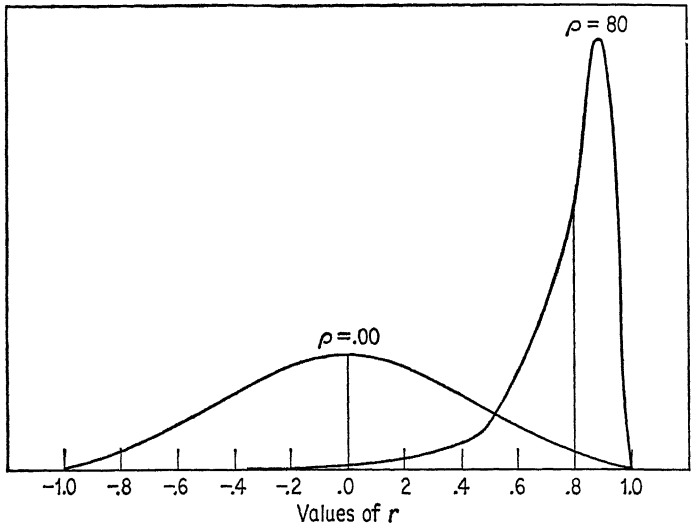


Figure 6.1 The sampling distribution of the correlation coefficient for random samples of $n = 8$ observations drawn from two populations having the indicated values of ρ.

the Appendix. It is possible to enter Table VII with a given value of r and to read directly the corresponding value of z'. For negative values of r the tabled values should be given a negative sign.

Fisher (1921) has shown that the distribution of z' is approximately normal and, for all practical purposes, may be considered independent of

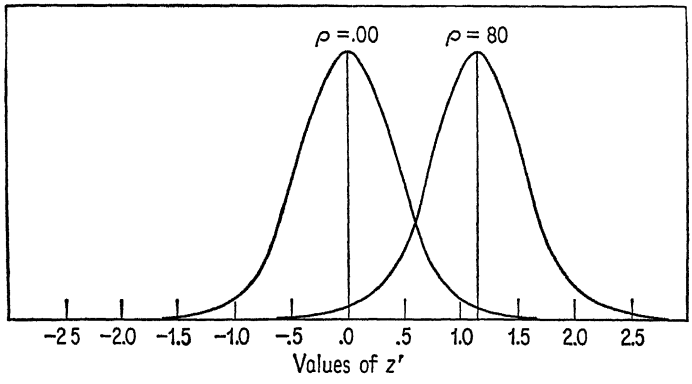


Figure 6.2 The sampling distribution of z' for random samples of $n = 8$ observations drawn from two populations having the indicated values of ρ.

the magnitude of the population correlation. Some indication of the extent to which the z' transformation normalizes the distribution of r for small samples may be gained from an examination of Figures 6.1 and 6.2. The first figure shows the distribution of r based upon $n = 8$ observations drawn from a population in which the population correlation is zero and also the distribution of r for $n = 8$ observations drawn from a population in which the population correlation is .80. The second figure shows the two distributions of z' for $n = 8$ observations drawn from the same populations. It is clear that the z' transformation has to a great degree normalized the sampling distribution of r.

Test of Significance

To illustrate the use of the z' transformation, let us suppose we have obtained a correlation coefficient of .82 based upon $n = 20$ observations. Previous research has shown that in the population of interest the population correlation is .85. The null hypothesis we wish to test is that we have a random sample from the same population, that is, from a population in which the population correlation is .85.

Fisher (1921) has also shown that the standard error of z' is given by

$$\sigma_{z'} = \frac{1}{\sqrt{n-3}} \tag{6.4}$$

In the present problem, we have $n = 20$, and therefore

$$\sigma_{z'} = \frac{1}{\sqrt{20-3}} = .243$$

Then, since z' is approximately normally distributed with known standard error, the test of the null hypothesis can be made by finding the normal deviate

$$z = \frac{z' - \bar{z}'}{\sigma_{z'}} \tag{6 5}$$

where $\bar{z}'$ is the mean of the sampling distribution of z' and corresponds to the value of the population correlation.

For the present problem we have $r = .82$ with $z' = 1.157$. If the population correlation is .85, then $\bar{z}' = 1.256$. Substituting in formula (6.5) we have

$$z = \frac{1.157 - 1.256}{.243} = -.407$$

By reference to the table of the normal curve, we find that an absolute value of z equal to 1.96 will be required for significance at the 5 per cent level with a two-sided test. Thus, the result of the test, with $\alpha = .05$,

provides no significant evidence against the null hypothesis that the sample has been drawn from a population in which the population correlation is .85.

Confidence Limits

By methods to be discussed in detail in the next chapter, it is possible to establish an interval such that we can say we have a certain degree of confidence that the interval contains the population correlation. The interval is called a *confidence interval* and the limits of the interval are called *confidence limits*. Applied to the correlation coefficient, the methods result in our first finding either

$$z' \pm 1.96\sigma_{z'} \tag{6.6}$$

for a 95 per cent confidence interval, or

$$z' \pm 2.58\sigma_{z'} \tag{6.7}$$

for a 99 per cent confidence interval. In the present problem, the 95 per cent confidence limits are obtained by finding

$$1.157 \pm (1.96)(.243)$$

or .681 and 1.633. From Table VII we find that the values of r corresponding to .681 and 1.633 are approximately .59 and .93 and these are the 95 per cent confidence limits for the example being considered.

If we say that we believe that the population correlation falls within the interval .59 to .93, and that we are 95 per cent confident that our belief is correct, we are expressing our degree of confidence that, in repeated sampling, such an inference concerning the population correlation will be correct 95 times in 100. For any particular sample, this inference will be either right or wrong; that is, either the population correlation falls within the interval or it does not.

SIGNIFICANCE OF THE DIFFERENCE BETWEEN TWO r's

The z' transformation may also be used in testing the significance of the difference between two correlation coefficients obtained from two independent samples of n_1 and n_2 observations. Assume, for example, we have obtained $r_1 = .73$ with $n_1 = 100$ observations and $r_2 = .60$ with $n_2 = 35$ observations. We wish to determine whether these two values of r differ significantly. The standard error of the difference between two independent z' values is given by

$$\sigma_{z_1'-z_2'} = \sqrt{\sigma_{z_1'}^2 + \sigma_{z_2'}^2}$$

or

$$\sigma_{z_1'-z_2'} = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}} \tag{6.8}$$

which, for the present example, gives us

$$\sigma_{z_1'-z_2'} = \sqrt{\frac{1}{100-3} + \frac{1}{35-3}}$$
$$= \sqrt{.010309 + .031250}$$
$$= .204$$

Then, since both z_1' and z_2' are approximately normally distributed, the difference between them will also be normally distributed about the population mean difference with standard error as given by formula (6.8). Thus we can obtain the normal deviate

$$z = \frac{(z_1' - z_2') - (\bar{z}_1' - \bar{z}_2')}{\sigma_{z_1'-z_2'}} \tag{6.9}$$

The null hypothesis we wish to test is that $\bar{z}_1' = \bar{z}_2'$ so that $\bar{z}_1' - \bar{z}_2' = 0$ We have $r_1 = .73$ with $z_1' = .929$ and $r_2 = .60$ with $z_2' = .693$. Then

$$z = \frac{.929 - .693}{.204} = 1.16$$

and, for a two-sided test with $\alpha = .05$, this is a nonsignificant value. We conclude that the two sample r's do not differ significantly.

TEST OF HOMOGENEITY OF k VALUES OF r

If we have more than $k = 2$ values of r and we wish to test the null hypothesis that they are homogeneous, that is, that the k values are all estimates of the same population value, we can make this test by finding

$$\chi^2 = \sum(n_i - 3)(z_i')^2 - \frac{[\sum(n_i - 3)(z_i')]^2}{\sum(n_i - 3)} \tag{6.10}$$

with $k - 1$ d.f. The calculations involved in the test for a set of $k = 4$ values of r are shown in Table 6.1. Substituting with the appropriate values from this table in formula (6.10), we have

$$\chi^2 = 73.126 - \frac{(109.140)^2}{168} = 2.224$$

a nonsignificant value for 3 d.f. We thus conclude that the various r's are homogeneous and can be considered as estimates of the same population value.

Since the r's given in Table 6.1 can be considered as estimates of the same population value, we find

$$\frac{\sum(n_i - 3)(z_i')}{\sum(n_i - 3)}$$

Table 6.1 Calculations for the χ^2 Test of Homogeneity of k Values of r

(1) Samples	(2) n_i	(3) r_i	(4) $n_i - 3$	(5) z_i'	(6) $(n_i - 3)(z_i')$	(7) $(n_i - 3)(z_i')^2$
1	33	.53	30	.590	17 700	10.443
2	58	.62	55	.725	39.875	28.909
3	42	.65	39	.775	30.225	23.424
4	47	.45	44	.485	21.340	10.350
$\sum$	180	2.25	168	2 575	109.140	73.126

as the weighted average value of z'. The weighted average in the present example is 109.140/168 = .650. From Table VII we find that the corresponding value of r is approximately .571, and we may regard this value as an estimate of the common population correlation based upon the data under consideration.

In combining r's from a fairly large number of samples by the method described above, a slight bias, present in each z' value, is accumulated which tends to make the estimate of the population value based upon the weighted z' values somewhat too large. If the χ^2 test does not result in rejection of the null hypothesis, then it may be desirable to obtain a more accurate estimate of the population correlation.

Fisher (1921) gives as a correction term for the bias

$$\frac{\rho}{2(n-1)}$$

which is to be subtracted from each z' value. The numerator of the correction term is the population correlation and this, of course, is unknown. However, the value of r corresponding to the average of the weighted z' values, obtained in the manner described, may be substituted for the population value in the correction term. The corrected z' for the first sample in Table 6.1 would then be

$$\text{Corrected } z' = .590 - \frac{.57}{2(33-1)} = .581$$

The corrected z' values for the other three samples are .720, .768, and .479, respectively. Weighting each corrected z' value by its degrees of freedom and summing, we obtain (30)(.581) + (55)(.720) + (39)(.768) + (44)(.479) = 108.058. Then we have 108.058/168 = .643, and the r corresponding to this z' value is approximately .567.

The correction term for the z' values is of relatively little importance when the n's of the various samples are large and when we have a small number of samples. With small n's and a large number of samples, the

correction may result in an estimate of the population correlation which differs considerably from that obtained from the uncorrected z' values.

SIGNIFICANCE OF THE DIFFERENCE BETWEEN NONINDEPENDENT r's

In some cases we may wish to test the significance of the difference between two correlation coefficients when the two values are not independent. For example, suppose we have measures on three variables, X_1, X_2, and Y for a group of n subjects. We denote the two correlations of X_1 and X_2 with Y by r_1 and r_2 and the correlation between X_1 and X_2 by r_{12}. To determine whether r_1 and r_2 differ significantly, we use a test due to Hotelling (1940). We find

$$t = (r_1 - r_2)\sqrt{\frac{(n-3)(1+r_{12})}{2(1 - r_1^2 - r_2^2 - r_{12}^2 + 2r_1r_2r_{12})}} \tag{6.11}$$

with $n - 3$ d.f. The t obtained with formula (6.11) can be evaluated for significance by reference to the table of t with $n - 3$ d.f.

QUESTIONS AND PROBLEMS

1. A correlation coefficient of .42 is based upon 45 observations. Can we conclude that this is significant at the 5 per cent level?

2. A correlation of .82 is obtained with a set of 39 observations. Establish 95 per cent confidence limits.

3. Twenty subjects were divided at random into two groups of 10 subjects each. One group was assigned to an experimental condition and the other group served as a control. Observations on two variables were obtained for each group. The correlation for the experimental group was .62 and for the control group the correlation was .73. Can we conclude that these two correlation coefficients differ significantly?

4. Suppose we have divided subjects at random into two groups with $n_1 = 33$ and $n_2 = 58$. The correlation coefficient between two variables for the first group is .53 and for the second group it is .62. Can we conclude that these two correlation coefficients differ significantly? Make the test first using the z' transformation for r. Then use the χ^2 test to determine whether the two values are homogeneous. You should find that $\chi^2 = z^2$.

5. About 1,500 students in the high schools of Seattle were given two forms of an attitude test. Samples of 100 papers were randomly drawn from the entire set of papers, and the correlation (reliability coefficient) was computed between scores on the two forms of the test for each sample. The obtained values for five samples were .87, .90, .82, .79, and .91. Can we conclude that these values are homogeneous?

7

THE t TEST FOR MEANS

INTRODUCTION

Assume that X is a continuous and normally distributed variable with population mean equal to m and standard deviation equal to σ. If random samples of n observations each are drawn from this population, the sampling distribution of the means of the samples will also be normal in form. Then the standard deviation of the sampling distribution, the standard error of the mean, will be given by

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{7.1}$$

The numerical value of the standard error of the mean is obviously related to the population standard deviation and the number of observations in the samples. If the samples are drawn from a population with $\sigma = 10.0$, and if the sample size $n = 25$, then $\sigma_{\bar{x}} = 2.0$. If $\sigma = 20.0$, and if $n = 25$, then $\sigma_{\bar{x}} = 4.0$. To reduce either of these standard errors by ½, it would be necessary to quadruple the sample size. With samples of $n = 100$ observations, and samples drawn from a population with $\sigma = 10.0$, the standard error of the mean would be 1.0, and for samples drawn from the population with $\sigma = 20.0$, the standard error would be 2.0.

If random samples are drawn from a population in which the distribution of X is skewed, the sampling distribution of the mean will not be normal in form if n is small, but the sampling distribution will approach normality as n becomes large. Similarly, if the population distribution is flat, approaching a rectangular distribution, the sampling distribution of the means of random samples will not be normal in form if n is small, but will tend toward normality as n becomes large.

We made use of these facts in our discussion of tests of significance involving random samples from a binomial population. When $P = Q$ the population is rectangular in form. In this case the sampling distribution of p, the mean of a sample from a binomial population, is symmetrical and, with n as small as 10, a good approximation of the probabilities associated with the sampling distribution of p was obtained by assuming that p was normally distributed. When $P \neq Q$, the binomial population is skewed

and so also is the sampling distribution of p. Yet, even with $P = \frac{1}{3}$, we found that the probabilities associated with the sampling distribution of p were approximated fairly well by assuming that p was normally distributed when nP was equal to or greater than 5. If we have a binomial population that is even more skewed, with, let us say, $P = \frac{1}{5}$, the probabilities associated with the sampling distribution of p would be fairly well approximated by assuming p to be normally distributed if $n = 25$, and the approximation would be even better if $n = 50$.

We know that the mean of the sampling distribution of p is equal to the mean of the binomial population, that is, $m = P$, and that the standard deviation of the population is $\sigma = \sqrt{PQ}$. Then $\sigma_p = \sigma/\sqrt{n} = \sqrt{PQ/n}$ and, assuming p to be normally distributed,

$$z = \frac{p - m}{\sigma_p}$$

is a normal deviate and can be evaluated by reference to the table of the unit normal curve. We were thus able to determine the probability associated with a particular sample value of p, assuming the sample had been drawn at random from a specified binomial population.

SAMPLING DISTRIBUTION OF THE MEAN

In this chapter we shall be concerned with a continuous variable X, which we shall assume to be normally distributed in the population. If we have a sample of n observations from this population, the sample mean will be designated by $\bar{X}$ and will be equal to

$$\bar{X} = \frac{\sum X}{n} \tag{7.2}$$

If we draw *random* samples of n observations each from the population, the sampling distribution of the means will be normal in form. The mean of the sampling distribution, as in the case of p, will be equal to the population mean. The population mean, although unknown, could be specified by hypothesis as we did in the case of P, the mean of the binomial population. For the binomial population, once P is specified, the population standard deviation, $\sigma = \sqrt{PQ}$, is also known. However, for our continuous variable X, specifying the population mean by hypothesis tells us nothing about the population standard deviation. This value still remains unknown. If it were known, then we could find the standard error of the mean, by formula (7.1), and

$$z = \frac{\bar{X} - m}{\sigma_{\bar{x}}}$$

would be a normal deviate and could be evaluated by means of the table of the unit normal curve.

We define the variance of the observations in a given sample of n observations as

$$s^2 = \frac{\sum(X - \bar{X})^2}{n - 1} \tag{7.3}$$

The variance, as defined above, is said to have $n - 1$ *degrees of freedom*, and is an estimate of the population variance σ^2. The standard deviation of the sample of n observations will be

$$s = \sqrt{\frac{\sum(X - \bar{X})^2}{n - 1}} \tag{7.4}$$

Then the standard error of the mean of a sample of n observations will be

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \tag{7.5}$$

and $s_{\bar{x}}$ is an estimate of $\sigma_{\bar{x}}$. The difference between $\bar{X}$, a sample mean, and m, a population mean, divided by $s_{\bar{x}}$, we shall define as t. Thus

$$t = \frac{\bar{X} - m}{s_{\bar{x}}} \tag{7.6}$$

with degrees of freedom equal to those associated with s^2 or $n - 1$.

THE t DISTRIBUTION

The distribution of t depends upon the number of degrees of freedom available in the set of n observations upon which s^2 is based. Hence, the table of t is a two-dimensional table which must be entered with both the obtained value of t from formula (7.6) and the number of degrees of freedom. The distribution of t is not normal for small samples. Its distribution is symmetrical, as is the distribution of z, but beyond a certain point (depending upon the number of degrees of freedom available) the curve of t does not approach the base line as rapidly as does the curve of the z distribution. This means, for example, that in order to cut off 5 per cent of the total area in the right tail of the t distribution, we shall have to go out from the mean beyond the value of $z = 1.65$ that cuts off 5 per cent of the total area in the right tail of the z distribution. Just how far out we shall have to go again depends upon the number of degrees of freedom available.

The value of s^2 itself is subject to sampling variation. As n increases, the accuracy with which s^2 estimates σ^2 increases also. For very large values

of n, the discrepancy between s^2 and σ^2 may be sufficiently small as to be negligible. Thus, in the limiting case, with n indefinitely large, the distribution of t is the same as the distribution of z. Table V, in the Appendix, is a table of the values of t significant at given levels of significance for varying degrees of freedom. It may be observed in the table that as the degrees of freedom increase, a smaller value of t is required for significance, until, for an infinite number of degrees of freedom, the values of t significant at the 5 and 1 per cent levels correspond to the significant values of z at the 5 and 1 per cent levels.

CONFIDENCE LIMITS FOR THE MEAN

Suppose we have a random sample of $n = 49$ observations, with $\bar{X} = 62.0$, and $s = 14.0$. The standard error of the mean, as given by formula (7.5), will then be $s_{\bar{x}} = 14.0/\sqrt{49} = 2.0$. Suppose also that we decide to reject any hypothesis concerning the population mean if we obtain, from formula (7.6), a t that is equal to or less than the $-t$ that cuts off .025 of the total area in the left tail of the t distribution or a t that is equal to or greater than the t that cuts off .025 of the total area in the right tail of the t distribution. Since the t distribution we are concerned with has $n - 1 = 49 - 1 = 48$ d.f., we find from the table of t that these two values will be -2.01 and 2.01, respectively. Then we may set up the following inequality

$$-t \leqq \frac{\bar{X} - m}{s_{\bar{x}}} \leqq t \tag{7.7}$$

Substituting in the above inequality with $t = 2.01$, $\bar{X} = 62.0$, $s_{\bar{x}} = 2.0$, and $-t = -2.01$, we have

$$-2.01 \leqq \frac{62.0 - m}{2.0} \leqq 2.01$$

or

$$(2.0)(-2.01) - 62.0 \leqq -m \leqq (2.0)(2.01) - 62.0$$

Multiplying by -1, remembering that the sense of an inequality is changed if the terms are multiplied by the same negative number, we obtain

$$62.0 + (2.0)(2.01) \geqq m \geqq 62.0 - (2.0)(2.01)$$

or

$$66.02 \geqq m \geqq 57.98$$

The interval 57.98 to 66.02 that we have just found is called a *confidence interval* and the limits of the interval are called *confidence limits*. The degree of confidence we have in the statement that m falls within the confidence interval is called a *confidence coefficient*. In the illustrative example, we have determined a 95 per cent confidence interval.

Confidence limits are statistics and like all statistics they are also subject to sampling variation. If we drew another sample of $n = 49$ observations from the same population as the first sample, both the sample mean and the standard deviation may be expected to be different from the values we obtained for the first sample. The 95 per cent confidence limits thus established for the second sample would not necessarily be the same as those established by the first sample. When we say we are 95 per cent confident that m falls within the 95 per cent confidence interval, we are expressing our degree of confidence that, in repeated sampling, such an inference concerning m will be correct 95 times in 100. For any particular sample, the inference will be right or wrong; that is, either m falls within the interval or it does not.

We may note that when we establish a confidence interval the procedure implies a test of significance. In essence, with $\alpha = .05$ and a two-tailed test of significance, we would reject, in the example being considered, any hypothesis that $m \leqq 57.98$ or that $m \geqq 66.02$.

With $n = 49$ and with $s = 14.0$, the 95 per cent confidence limits are 57.98 and 66.02. Increasing n to 100 observations, that is, slightly more than doubling the sample size, will serve to reduce the confidence interval in two ways, assuming that s^2, our estimate of the population variance, remains the same.

In the first place, the standard error of the mean will now be $s_{\bar{x}} = 14.0/\sqrt{100} = 1.4$, as compared with the value of 2.0 when the sample consisted of only 49 observations. In the second place, the values of t cutting off .025 of the total area in the two tails of the t distribution for 99 d.f. are -1.984 and 1.984 rather than the values of -2.01 and 2.01 for 48 d.f. We would thus have as the 95 per cent confidence interval

$$62.0 + (1.4)(1.984) \geqq m \geqq 62.0 - (1.4)(1.984)$$

$$64.78 \geqq m \geqq 59.22$$

This 95 per cent confidence interval, based upon $n = 100$ observations, has a range of $64.78 - 59.22 = 5.56$, whereas that based upon $n = 49$ observations had a range of $66.02 - 57.98 = 8.04$. It should be clear that, if we wish a narrow confidence interval, we shall need to make a large number of observations when the estimated standard deviation of the population is as large as 14.0.

DIFFERENCE BETWEEN TWO MEANS

In an experiment upon the influence of two treatments upon retention, the treatments were assigned at random in such a way that 20 subjects received Treatment 1 (T_1) and 20 subjects received Treatment 2 (T_2).

Subjects in both groups were presented with a series of paired words and were asked to guess which word in each pair was "correct." T_1 consisted of giving each subject a slight shock for each wrong guess. In T_2 the subjects were not shocked; instead each wrong guess was followed by the flashing of a red light. Subjects in both groups were trained to a criterion set by the experimenter and then retested after 24 hours. The dependent variable X is the number of correct responses made on the delayed test. The "retention scores" of the two groups are given in Table 7.1.

Table 7.1 Retention Scores for 20 Subjects Assigned to Treatment 1 and 20 Subjects Assigned to Treatment 2

Treatments	Scores				$\sum X$ and $\sum X^2$ for Each Treatment
T_1	12	16	6	10	$\sum X_1 = 220$
	6	13	16	12	
	7	14	13	11	$\sum X_1^2 = 2{,}596$
	12	9	10	9	
	10	14	7	13	
T_2	4	9	1	8	$\sum X_2 = 160$
	12	11	8	9	
	9	0	10	9	$\sum X_2^2 = 1{,}522$
	9	9	8	10	
	14	11	6	3	

The mean score for T_1 is

$$\bar{X}_1 = \frac{220}{20} = 11.0$$

and the mean score for T_2 is

$$\bar{X}_2 = \frac{160}{20} = 8.0$$

The difference between these two means is $\bar{X}_1 - \bar{X}_2 = 11.0 - 8.0 = 3.0$. If the experiment were repeated under the same conditions an indefinitely large number of times, we would not expect to obtain exactly the same values of $\bar{X}_1$ and $\bar{X}_2$ in these repetitions that we obtained in the particular experiment under consideration. The means of both samples are subject to random sampling variation and this will also be true of the difference between the means. However, the sampling distribution of $\bar{X}_1$ will be normally distributed about the population mean m_1 and the sampling distribution of $\bar{X}_2$ will be normally distributed about the population mean m_2. The sampling distribution of the difference, $\bar{X}_1 - \bar{X}_2$, will also be normally distributed about the population mean difference $m_1 - m_2$. If we knew the standard error of the sampling distribution of $\bar{X}_1 - \bar{X}_2$, it would be possible

to evaluate the particular difference obtained in this experiment by means of the t distribution.

STANDARD ERROR OF THE DIFFERENCE BETWEEN TWO MEANS

The standard error of the difference between the means of two independent random samples will be given by

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2} \tag{7.8}$$

and this may also be written as

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{7.9}$$

In the present problem σ_1^2 and σ_2^2 are unknown, but each may be estimated by means of formula (7.3). Thus the estimated standard error of the difference between the means will be

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{7.10}$$

For the moment, let us assume that $\sigma_1^2 = \sigma_2^2$ so that s_1^2 and s_2^2 are both estimates of the same population variance. If we have two or more estimates of a common parameter, these may be combined in such a way as to yield a single estimate. In the case of k sample variances, all of which are assumed to estimate the same common population variance, the single estimate is obtained by

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2}{(n_1 - 1) + (n_2 - 1) + \cdots + (n_k - 1)} \tag{7.11}$$

with d.f. equal to $\sum n - k$. If all the n's are equal, then the d.f. will be equal to $k(n - 1)$ where n is the number of observations in each sample. We let

$$\sum x^2 = \sum (X - \bar{X})^2 \tag{7.12}$$

or the sum of squared deviations of the n observations in a given sample from the sample mean. Then we also have $\sum x^2 = (n - 1)s^2$ and

$$s^2 = \frac{\sum x_1^2 + \sum x_2^2 + \cdots + \sum x_k^2}{\sum n - k} \tag{7.13}$$

For the present problem, we have $k = 2$ and therefore

$$s^2 = \frac{\sum x_1^2 + \sum x_2^2}{n_1 + n_2 - 2} \tag{7.14}$$

with d.f. equal to $n_1 + n_2 - 2$. Substituting with the single estimate s^2 for the separate estimates ${s_1}^2$ and ${s_2}^2$ in formula (7.10), we have

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\frac{\sum {x_1}^2 + \sum {x_2}^2}{n_1 + n_2 - 2}}{n_1} + \frac{\frac{\sum {x_1}^2 + \sum {x_2}^2}{n_1 + n_2 - 2}}{n_2}}$$

and this may be written

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\left(\frac{\sum {x_1}^2 + \sum {x_2}^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \tag{7.15}$$

We observe also that if $n_1 = n_2 = n$, then

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{2s^2}{n}} = s\sqrt{\frac{2}{n}} \tag{7.16}$$

where n is the number of observations in each group.

It should be emphasized that $\sum {x_1}^2$ refers to the sum of squared deviations of the n_1 observations, obtained under T_1, about the mean for T_1, and similarly $\sum {x_2}^2$ refers to the sum of the squared deviations of the n_2 observations, obtained under T_2, about the mean for T_2. A convenient method for calculating these sums of squares is

$$\sum x^2 = \sum X^2 - \frac{(\sum X)^2}{n} \tag{7.17}$$

By formula (7.17), we find that the sum of squares for T_1 is

$$\sum {x_1}^2 = 2{,}596 - \frac{(220)^2}{20} = 176.00$$

and the sum of squares for T_2 is

$$\sum {x_2}^2 = 1{,}522 - \frac{(160)^2}{20} = 242.00$$

Substituting in formula (7.15) for the standard error of the difference between the means, we have

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\left(\frac{176 + 242}{20 + 20 - 2}\right)\left(\frac{1}{20} + \frac{1}{20}\right)} = \sqrt{\left(\frac{418}{38}\right)\left(\frac{2}{20}\right)} = 1.049$$

CONFIDENCE LIMITS FOR A MEAN DIFFERENCE

We can, in the manner described previously, find 95 or 99 per cent confidence limits for the population mean difference, $m_1 - m_2$. With

$n_1 + n_2 - 2 = 20 + 20 - 2 = 38$ d.f., we find that a t of -2.711 will cut off .005 of the total area in the left tail and a t of 2.711 will cut off .005 of the total area in the right tail of the t distribution. Then the 99 per cent confidence limits will be given by

$$-2.711 \leqq \frac{(11.0 - 8.0) - (m_1 - m_2)}{1.049} \leqq 2.711$$

or

$$3.0 + (1.049)(2.711) \geqq (m_1 - m_2) \geqq 3.0 - (1.049)(2.711)$$

$$5.84 \geqq (m_1 - m_2) \geqq .16$$

and we can say that we are 99 per cent confident that the population mean difference, $m_1 - m_2$, is within these limits.

TEST OF SIGNIFICANCE OF A MEAN DIFFERENCE

If our major interest is in determining whether a specified null hypothesis concerning $m_1 - m_2$ is to be rejected, then this hypothesis may be tested by finding

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (m_1 - m_2)}{s_{\bar{x}_1 - \bar{x}_2}} \tag{7.18}$$

Specifically, if the null hypothesis is $m_1 = m_2$, so that $m_1 - m_2 = 0$, then

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{x}_1 - \bar{x}_2}} \tag{7.19}$$

and for the present problem we have

$$t = \frac{11.0 - 8.0}{1.049} = 2.86$$

with 38 d.f.

THE NULL HYPOTHESIS AND ALTERNATIVES

Before asking whether or not the value of t obtained above is significant, we should consider more specifically the nature of the test of significance. In general, we test a null hypothesis against a class of alternative hypotheses. If the null hypothesis is false, so that one of the alternative hypotheses is true, we can define the power of a test of significance as

$$\text{Power} = 1 - \text{Probability of a Type II error}$$

Since the probability of a Type II error is the probability of *not* rejecting the null hypothesis when it is false, the power of a test of significance can

be said to be the probability of rejecting the null hypothesis when it *should* be rejected.

One way in which we can increase the power of a given test is to make α large. But we do not like to make α too large, since by doing so we increase the probability of a Type I error. If we hold α constant, then we can also increase the power of a given test by increasing the number of observations in the sample under consideration. If we hold both α and the number of observations in the sample constant, then we can increase the power of a test against a *selected* class of alternatives to the null hypothesis by the manner in which we choose the critical region of rejection in the t distribution. It is this latter manner of increasing the power of a test that we now consider.

If we designate the null hypothesis as H_0 and the alternatives to this hypothesis as H_1, then we may be interested in any one of the following three tests:

Test 1	$H_0: m_1 = m_2$	with	$H_1: m_1 \neq m_2$
Test 2	$H_0: m_1 \leqq m_2$	with	$H_1: m_1 > m_2$
Test 3	$H_0: m_1 \geqq m_2$	with	$H_1: m_1 < m_2$

Suppose we choose $\alpha = .05$. If we make Test 1, we shall reject the null hypothesis if the t we obtain falls in either of the two shaded areas of Figure 7.1. With 38 d.f., the critical values of t, those that would result in the rejection of the null hypothesis, would be any t equal to or less than -2.025 or any t equal to or greater than 2.025. These are the values of t cutting off .025 of the total area in each tail of the distribution. Since the areas of rejection for Test 1 are in either one of the two tails of the t distribution, this test is called a *two-tailed* test or *two-sided* test. Test 1 provides protection against the possibility of $m_1 > m_2$ and also the possibility of

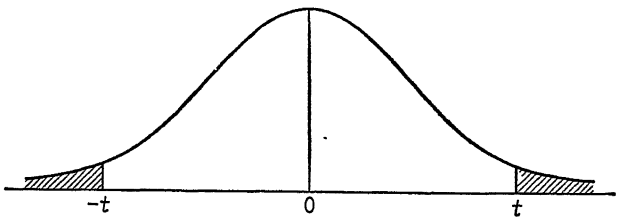


Figure 7.1 The two-tailed test of significance of the null hypothesis $m_1 = m_2$ against the alternative $m_1 \neq m_2$. Each of the shaded areas in the two tails of the t distribution is .025 of the total area. With $\alpha = .05$, the null hypothesis is rejected if the observed value of t falls in either of the two shaded areas.

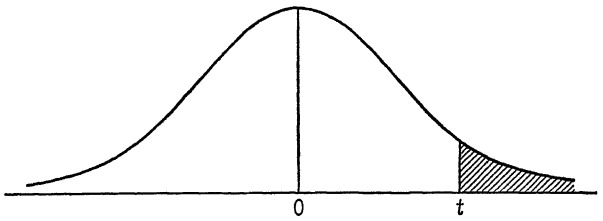


Figure 7.2 The one-tailed test of significance of the null hypothesis $m_1 \leqq m_2$ against the alternative $m_1 > m_2$. The shaded area in the right tail of the t distribution is .05 of the total area. With $\alpha = .05$, the null hypothesis is rejected if the observed value of t falls in the shaded area.

$m_1 < m_2$. In other words, it is sensitive to the absolute value of the difference between m_1 and m_2. Test 1 is the one we should use if we are interested in the absolute magnitude of the difference between the means, $m_1 - m_2$, and not specifically in the direction of the difference.

If we make Test 2, then we shall reject the null hypothesis only if the obtained value of t falls in the shaded area of Figure 7.2. If $\alpha = .05$, then we want the area in the right tail to correspond to .05 of the total area of the t distribution. With 38 d.f., the critical value of t, cutting off .05 in the right tail, is approximately 1.68. Since the area of rejection is the right tail of the t distribution, Test 2 is referred to as a *right-tailed*, a *one-tailed*, or a *one-sided* test. Test 2 provides protection against the class of alternatives $m_1 > m_2$ only. If it is true that $m_1 > m_2$, then Test 2 will be more powerful than Test 1 against *this class of alternatives*, but, unlike Test 1, Test 2 provides no protection against the possibility of $m_1 < m_2$. Test 2 should be used when we wish to reject the null hypothesis *only* if $m_1 > m_2$ and we have no interest in the possibility that $m_1 < m_2$.

With Test 3, the region of rejection of the null hypothesis is the left tail of the t distribution, as shown in Figure 7.3. If $\alpha = .05$ and with 38 d.f., then for Test 3 the critical value of t is approximately -1.68. Test 3, like Test 2, is a one-tailed or one-sided test. Test 3 provides protection only against the class of alternatives $m_1 < m_2$. If one of the alternatives is true, then Test 3 will be more powerful than Test 1 against *this* alternative, but will not protect against the possibility of $m_1 > m_2$. Test 3 should be used, therefore, when we wish to reject the null hypothesis *only* if $m_1 < m_2$ and we have no interest in the possibility that $m_1 > m_2$.

Let us assume that, in planning the experiment described above, the experimenter decided he was interested in the difference between the means without regard to the direction of the difference Then, making a two-sided test (Test 1), he has $t = 2.86$ with 38 d.f. and this is a significant value.

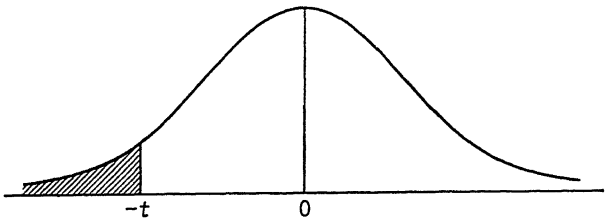


Figure 7.3 The one-tailed test of significance of the null hypothesis $m_1 \geq m_2$ against the alternative $m_1 < m_2$. The shaded area in the left tail of the t distribution is .05 of the total area. With $\alpha = .05$, the null hypothesis is rejected if the observed value of t falls in the shaded area.

It must be emphasized that if Test 2 or Test 3 is to be made in a given experiment, this decision *must* be made at the time the experiment is planned and should not be suggested by an examination of the results of the experiment. It may sometimes happen that the difference between two means in an experiment will be declared significant if a one-tailed test is made, but nonsignificant if a two-tailed test is made To decide, after looking at the data, that a one-tailed test is to be made is not only unscientific; it is also dishonest.

We have previously discussed some experiments involving a legitimate use of a one-tailed test. In taste discrimination experiments, for example, we are ordinarily interested in alternatives to the null hypothesis that indicate a better than chance ability to make correct discriminations. As another example of the appropriate use of a one-tailed test, consider the case of the farmer from Whidbey Island. The null hypothesis we tested was $P = \frac{1}{2}$ with the alternative being $P > \frac{1}{2}$ In using the normal curve to evaluate the results of this experiment, the region of rejection was the right tail of the curve, that is, we made a one-tailed test corresponding to Test 2 above. This particular test was made because we were interested only in the possibility that the farmer could do better than chance and we had no interest in the possibility that he might do worse than chance.

NUMBER OF OBSERVATIONS

A problem faced in many experiments is a decision regarding the number of observations to be made for each treatment. Assume that on the basis of previous experience we know something about the variability of the observations under the treatments. It will, in fact, simplify the presentation if we can assume that the common population variance, σ^2, is known, so that we can make use of the table of the unit normal curve rather than the t table. Assume $\sigma = 10$. Suppose also that we set $\alpha = .05$

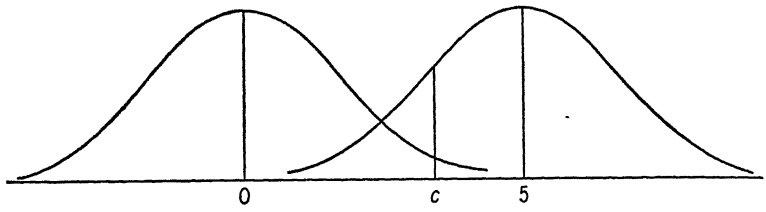


Figure 7.4 The sampling distribution of $\bar{X}_1 - \bar{X}_2$ when $m_1 - m_2 = 5$, at the right, and the sampling distribution of $\bar{X}_1 - \bar{X}_2$ when $m_1 - m_2 = 0$, at the left. The point c is located so as to cut off .025 of the total area in the right tail of the distribution at the left and .16 of the total area in the left tail of the distribution at the right.

and we have decided upon a two-tailed test. For a two-tailed test, with $\alpha = .05$, the critical values of z are 1.96 and -1.96. Suppose we also decide that the population difference, $m_1 - m_2$, must be equal to or greater than 5 or equal to or less than -5 to be of theoretical or practical interest. Furthermore, we want the probability of a Type II error, if the true difference is 5 or -5, to be no greater than .16. We have previously defined the power of a test as 1 minus the probability of a Type II error. We desire the test to have a power, therefore, of $1 - .16 = .84$, which is to say that we want the test to have a probability of at least .84 of rejecting the null hypothesis if either one of the alternatives is true.

Consider first only the alternative that $m_1 - m_2 = 5$. Figure 7.4 shows the sampling distribution of $\bar{X}_1 - \bar{X}_2$, when $m_1 - m_2 = 5$, at the right, and the sampling distribution of $\bar{X}_1 - \bar{X}_2$, when $m_1 - m_2 = 0$, at the left. Let $z_0 = 1.96$ be the critical value of z resulting in the rejection of the null hypothesis when it is true, that is, when $m_1 - m_2 = 0$. Then, if the null hypothesis is true, the value of c in Figure 7.4 will be

$$c = 0 + z_0\sigma_{\bar{x}_1-\bar{x}_2}$$
$$= 0 + (1.96)\sigma_{\bar{x}_1-\bar{x}_2}$$

If the null hypothesis is false and $m_1 - m_2 = 5$, and if we obtain a difference that falls to the left of c, the null hypothesis will not be rejected and we shall make a Type II error. We want this probability to be no greater than .16. Thus, for the sampling distribution when $m_1 - m_2 = 5$, we want .16 of the area to fall to the left of c and .84 to the right. Let the z value corresponding to c, in this instance, be z_1 and from the table of the normal curve we find $z_1 = -1.00$. Then we also have

$$c = 5 + z_1\sigma_{\bar{x}_1-\bar{x}_2}$$
$$= 5 + (-1.00)\sigma_{\bar{x}_1-\bar{x}_2}$$

Using the two equations for c, we have

$$0 + (1.96)\sigma_{\bar{x}_1-\bar{x}_2} = 5 - (1.00)\sigma_{\bar{x}_1-\bar{x}_2}$$

or

$$\begin{aligned} 5 &= (1.96)\sigma_{\bar{x}_1-\bar{x}_2} + (1.00)\sigma_{\bar{x}_1-\bar{x}_2} \\ &= \sigma_{\bar{x}_1-\bar{x}_2}(1.96 + 1.00) \end{aligned}$$

With $n_1 = n_2 = n$, we have $\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{2\sigma^2/n}$ and, therefore,

$$5 = \sqrt{\frac{2\sigma^2}{n}}\,(1.96 + 1.00)$$

or

$$5^2 = \frac{2\sigma^2}{n}\,(1.96 + 1.00)^2$$

Solving for n, with $\sigma = 10$, we have

$$\begin{aligned} n &= \frac{2(10)^2}{(5)^2}\,(2.96)^2 \\ &= 70 \end{aligned}$$

and we would want to have $n = 70$ subjects in each group, if $\sigma = 10$. We would obtain the same result by considering the alternative $m_1 - m_2 = -5$, with $\sigma = 10$.

We shall not lose much in the way of accuracy if we take $z_0 = 2.00$, rather than 1.96. Then, for the two-tailed test, if we want the probability of a Type II error to be no greater than .16, we have the general formula

$$n = \frac{2\sigma^2}{(m_1 - m_2)^2}\,(2.00 + 1.00)^2$$

or

$$n = 18\,\frac{\sigma^2}{(m_1 - m_2)^2} \tag{7.20}$$

for the required number of subjects in each group.

If we are willing to risk a greater probability of a Type II error, say, .50, then for the two-tailed test with $\alpha = .05$, we have

$$\begin{aligned} n &= \frac{2\sigma^2}{(m_1 - m_2)^2}\,(2.00 + 0)^2 \\ &= 8\,\frac{\sigma^2}{(m_1 - m_2)^2} \end{aligned} \tag{7.21}$$

for the required number of subjects in each group.

a number corresponding to a column. These three numbers will give a point of entry into the table. Once we have the point of entry, it makes no difference whether we read up, down, or across the table. Let us suppose the point of entry is 02, 02, and 05.

Suppose we have 100 subjects and we wish to select, at random, two groups of 10 subjects each. We assign the numbers 00, 01, 02, ···, 99 to the 100 subjects. It does not matter which subject receives which number; it is only necessary that each subject have a different number. Since the numbers assigned to the subjects consist of two-digit numbers, we shall make use of columns 05 and 06 read downward. We read down the columns at the point of entry, selecting the first 20 unlike numbers in the set 00 to 99. The first few numbers we encounter are 52, 98, 55, 94, 87, 42, and 30. We continue reading until we have 20 unlike numbers corresponding to 20 of the 100 subjects. The first 10 subjects selected in this way will be assigned to Treatment 1 and the second 10 to Treatment 2.

QUESTIONS AND PROBLEMS

1. Given a random sample of 16 cases with mean equal to 22.4 and s equal to 4.3. Establish 95 per cent confidence limits.

2. The mean score on a standardized test for a random sample of 200 freshmen college students at University A is 133.8, with s equal to 14.7. For a random sample of 140 freshmen at University B, the mean score is 138.4, with s equal to 15.2. Determine whether the two means differ significantly.

3. Forty subjects are assigned at random to two treatments, with 20 subjects for each treatment. The measures on the dependent variable are given below:

Treatment 1				Treatment 2			
39	41	39	44	36	41	30	39
39	40	39	40	36	39	33	37
37	42	37	43	35	42	36	37
44	38	38	38	34	38	33	31
43	38	41	39	40	32	33	38

Determine whether the two treatment means differ significantly. If an automatic calculating machine is not available, the calculations may be somewhat easier if a constant, say 30, is subtracted from each measure. If the same constant is subtracted from each measure, this will not influence the difference between the means, nor will it change the variance.

4. Morgan (1945) designed an experiment to test the hypothesis that failure to solve a problem tends to foster inductive reasoning more than immediate success. "*S*'s were confronted with the problem of discovering which of six cues to follow in order to make a bell ring. In one group (called the restricted hypothesis group) the cue which would make the bell ring was predetermined by the *E*. In another group (called the unrestricted hypothesis group) success followed the use of any cue by the *S*. Interspersed throughout the experiment

were test series to determine how well the *S*'s in both groups could discover a predetermined cue" (p. 146). The question is whether the restricted hypothesis group profited by the mistakes made in searching for the correct cue and surpassed, on the test series, the performance of the subjects in the unrestricted group. The data are as follows:

Unrestricted Group					Restricted Group				
6	12	14	19	35	4	8	9	12	25
7	12	14	23		5	8	9	13	
8	12	15	24		6	8	10	13	
10	13	15	30		6	9	10	15	
10	14	16	34		7	9	11	15	

Determine whether the two treatment means differ significantly.

5. In an experiment the sum of squared deviations for one treatment group was 420 and for the other treatment group the sum of squared deviations was 482. Each group had $n = 25$ subjects. The difference between the treatment means was 3.03. Is this difference significant at the 5 per cent level?

6. Measures obtained on a dependent variable for two treatment groups are given below:

Treatment 1				Treatment 2			
52	171	151	45	71	86	218	165
75	54	101		95	141	152	
170	104	74		151	52	120	
30	81	146		53	108	115	

Determine whether the difference between the treatment means is significant at the 5 per cent level.

7. In an experiment, 20 rats were randomly assigned to each of two conditions. The experimental condition consisted of giving each rat a 12-hour period of exploration in a maze. The other group served as a control group and was not given a period of exploration. Both groups were deprived of food for the same length of time and tested in the maze. Records were kept of the number of trials required to learn the maze to a criterion of one run with no errors. Data for the two groups are given below:

Control				Experimental			
10	7	9	6	12	7	9	6
8	6	10	13	5	9	9	9
9	7	12	12	6	4	8	4
15	6	9	11	9	10	11	6
9	13	4	9	9	7	10	7

Determine whether the difference between the means for the control and experimental groups is significant at the 5 per cent level.

8. In an experiment the standard error of the difference between two means was 1.42 with $n = 10$ subjects in each treatment group. A repetition of this experiment is planned and the experimenter wishes to be able to reject the null

hypothesis if the absolute difference between the population means is 2.56 or greater. On the basis of the data available, it is possible to solve for s^2. Assume that s^2 is the population variance. (*a*) How many subjects should the experimenter have in each group, if $\alpha = .05$ and if the probability of a Type II error is to be no greater than .16? (*b*) How many subjects should the experimenter have in each group if $\alpha = .05$ and if the probability of a Type II error is to be no greater than .50? In answering these two questions use the approximations given by formulas (7.20) and (7.21).

9. Give a brief interpretation of the meaning of confidence limits.

10. Comment upon the following statement: Establishing confidence limits always implies a test of significance

11. Discuss briefly the *t* test of a null hypothesis concerning a difference between two means in relation to the alternatives to the null hypothesis. (*a*) Under what conditions should we make a two-sided test? (*b*) Under what conditions should we make a right-tailed test? (*c*) Under what conditions should we make a left-tailed test? (*d*) Give an example where each test would be appropriate.

12. Define, briefly, each of the following terms:

confidence coefficient	one-sided test
confidence interval	power of a test
confidence limits	two-sided test

8

HETEROGENEITY OF VARIANCE AND THE *t* TEST

INTRODUCTION

The methods used in determining the standard error of the difference between two means and the t test used in evaluating the difference between the means, as described in the last chapter, are based on the assumption that the separate variance estimates provided by the two samples are both estimates of the same population variance. However, it may sometimes happen that one treatment will serve to increase or decrease the variability of the observations, whereas the other treatment may not. If the difference between the two treatment variances is significant, then the methods to be described in this chapter will be more appropriate than those of the previous chapter.

If the two sample variances, s_1^2 and s_2^2, are not equal, then we may make a test to determine whether the difference between them is statistically significant. Before describing the test of significance and the procedures to be used if the variances are found to differ significantly, let us consider some convenient guideposts. If we choose $\alpha = .05$, and if we have $n = 10$ observations in each sample, then one of the two variances will have to be approximately 4 times as large as the other in order for the difference between them to be significant. With $n = 20$ observations in each group, and with $\alpha = .05$, then one of the two variances will have to be approximately 2.5 times as large as the other for the difference between them to be significant. With 30 observations in each sample, then if one variance is approximately 2 times as large as the other, the difference between them will be significant at the 5 per cent level. As the number of observations in each sample is increased, smaller differences between the two variances will become significant.

THE *F* DISTRIBUTION

Given two sample variances, s_1^2 and s_2^2, we may test the null hypothesis $\sigma_1^2 = \sigma_2^2$ against the alternative hypothesis $\sigma_1^2 \neq \sigma_2^2$. The ratio

of the two sample variances is distributed in a manner discovered by Fisher (1936) and the significant values of the ratio at the .05 and .01 levels of significance have been calculated by Snedecor (1956), who named the ratio F in Fisher's honor. The significant values of F at the .25, .10, .025, and .005 levels of significance have been calculated by Merrington and Thompson (1943). Now if F is defined as

$$F = \frac{s_1^2}{s_2^2} \quad \text{or} \quad F = \frac{s_2^2}{s_1^2} \tag{8.1}$$

or as the ratio of two variances, then whether F will be greater than 1.00 or smaller than 1.00 will depend merely upon whether s_1^2 or s_2^2 is put in the numerator of the ratio. The tabled values of F, Table VIII and Table IX in the Appendix, are for a one-sided or one-tailed test and correspond to the probability of F greater than 1.00, when the null hypothesis is true.[1] Thus, to use the tables, we shall always find the value of F greater than 1.00 for formula (8.1) and this means we shall always put the larger of the two sample variances in the numerator.

If the alternative to the null hypothesis is $\sigma_1^2 \neq \sigma_2^2$—and in experimental work it usually is—then to protect against this alternative we need to make a two-sided or two-tailed test; that is, we want to reject the null hypothesis if either $\sigma_1^2 > \sigma_2^2$ or if $\sigma_1^2 < \sigma_2^2$. For the two-sided test, with $\alpha = .05$, the critical value of F will be the tabled value with probability .025. Similarly, for the two-sided test with $\alpha = .01$, the critical value of F will be the tabled value with probability .005.

TESTING FOR HOMOGENEITY OF VARIANCE

In the experiment on retention, described in the previous chapter, we had for Treatment 1, $\sum x_1^2 = 176$, and for Treatment 2, $\sum x_2^2 = 242$. Then

$$s_1^2 = \frac{176}{19} = 9.263 \quad \text{and} \quad s_2^2 = \frac{242}{19} = 12.737$$

Since s_2^2 is larger than s_1^2, we have

$$F = \frac{12.737}{9.263} = 1.375$$

To determine whether $F = 1.375$ is significant, we enter the column of Table IX with the degrees of freedom corresponding to the numerator

[1] Table VIII is Snedecor's table and gives the 5 and 1 per cent points for the distribution of F. Table IX is the Merrington and Thompson table and gives the 25, 10, 2.5, and 0.5 per cent points for the distribution of F.

of the F ratio and the row with the degrees of freedom corresponding to the denominator. For our obtained $F = 1\,375$, we have 19 d.f. for the numerator and 19 d.f. for the denominator. Table IX has no column corresponding to 19 d.f., but we find that the critical value of F, with $\alpha = .05$, for 20 and 19 d.f. is 2 51. Our obtained value of $F = 1.375$ is less than this critical value and, with $\alpha = .05$, the null hypothesis would not be rejected. Thus, the methods we used in the previous chapter to find the standard error of the difference and to evaluate the difference between the means in the experiment were appropriate.

The F test of formula (8.1) is often referred to as a test for *homogeneity of variance.* If a nonsignificant value is obtained, the two sample variances are said to be homogeneous, that is, they are both assumed to be estimates of the same population variance. With a significant value of F, the variances would be said to be heterogeneous.

HETEROGENEITY OF VARIANCE WITH $n_1 \neq n_2$

Let us suppose, in an experiment comparing two treatments, we do not have $n_1 = n_2 = n$, and that the two sample variances are heterogeneous. Consider, for example, the data of Table 8.1. Testing for homogeneity of

Table 8.1 Means and Variances for Two Treatments with Unequal n's

Treatment 1	Treatment 2
$\bar{X}_1 = 20.6$	$\bar{X}_2 = 16.0$
$s_1^2 = 28\,42$	$s_2^2 = 6.72$
$n_1 = 10$	$n_2 = 20$

variance, we have, since s_1^2 is greater than s_2^2,

$$F = \frac{28.42}{6.72} = 4.23$$

with 9 and 19 d.f. With $\alpha = .05$, the critical value of F is 2.88. Since our obtained value of $F = 4.23$ exceeds the critical value, we reject the null hypothesis.

Formula (7.14) assumes that the sample variances are estimates of the same population variance and combines the separate estimates in such a way as to provide a single estimate of the common population variance. But we have rejected the null hypothesis $\sigma_1^2 = \sigma_2^2$. Therefore, s_1^2 and s_2^2 cannot be said to be estimates of the same population variance and formula (7.14) is inappropriate for the case under consideration. Instead of using a single estimate, s^2, based upon formula (7.14), to find the standard error

of the difference between the two means, we shall use the separate estimates, s_1^2 and s_2^2. Then, by formula (7.10), we have

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{28.42}{10} + \frac{6.72}{20}} = 1.783$$

and

$$t = \frac{20.6 - 16.0}{1.783} = 2.58$$

To determine whether $t = 2.58$ is significant, we first find, from Table V, the critical values of t for $n_1 - 1 = 9$ d.f. and for $n_2 - 1 = 19$ d.f. For a two-sided test, with $\alpha = .05$, these two values are $t_1 = 2.262$ and $t_2 = 2.093$, respectively. Then we find

$$t_{.05} = \frac{t_1 \frac{s_1^2}{n_1} + t_2 \frac{s_2^2}{n_2}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{8.2}$$

The value of $t_{.05}$ obtained from formula (8.2) will be the critical value of t in terms of which our obtained $t = 2.58$ is to be evaluated.[2]

Substituting in formula (8.2), we have

$$t_{.05} = \frac{2.262 \frac{28.42}{10} + 2.093 \frac{6.72}{20}}{\frac{28.42}{10} + \frac{6.72}{20}} = \frac{7.132}{3.178} = 2.24$$

Since our obtained $t = 2.58$ exceeds $t_{.05} = 2.24$, the null hypothesis $m_1 = m_2$ will be rejected. Since we have obtained both a significant value of F and a significant value of t, we conclude that the two treatments have resulted in a significant difference in the treatment variances and also in the treatment means.

HETEROGENEITY OF VARIANCE WITH $n_1 = n_2$

For the variance of the difference between two means, assuming homogeneity of variance, we have the square of formula (7.15) or

$$s_{\bar{x}_1-\bar{x}_2}^2 = \frac{\sum x_1^2 + \sum x_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \tag{8.3}$$

[2] Formula (8.2) is an approximation developed by Cochran.

If we have $n_1 = n_2 = n$, then formula (8.3) can be written as

$$\left(\frac{\sum x_1^2 + \sum x_2^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \left(\frac{\sum x_1^2 + \sum x_2^2}{2n - 2}\right)\left(\frac{2}{n}\right) = \frac{\sum x_1^2 + \sum x_2^2}{n(n-1)}$$

For the variance of the difference between two means, with heterogeneity of variance, we have the square of formula (7.10) or

$$s_{\bar{x}_1 - \bar{x}_2}^2 = \frac{\frac{\sum x_1^2}{n_1 - 1}}{n_1} + \frac{\frac{\sum x_2^2}{n_2 - 1}}{n_2} \tag{8.4}$$

If $n_1 = n_2 = n$, then we can also write formula (8.4) as

$$\frac{\frac{\sum x_1^2}{n_1 - 1}}{n_1} + \frac{\frac{\sum x_2^2}{n_2 - 1}}{n_2} = \frac{\sum x_1^2 + \sum x_2^2}{n(n-1)}$$

We thus see that when $n_1 = n_2 = n$, formulas (8.3) and (8.4) are identical. Therefore, if $n_1 = n_2 = n$, we will obtain the same standard error of the difference between the two means, regardless of whether we use formula (8.3) with homogeneity of the sample variances or formula (8.4) with heterogeneity of the sample variances. The critical value for evaluating t, however, will depend upon whether or not we have homogeneity of variance. With homogeneity of variance, the critical value of t will be the tabled value for $n_1 + n_2 - 2 = 2n - 2$ d.f. On the other hand, if we have heterogeneity of variance, then the obtained value of t must be evaluated in terms of $t_{.05}$ of formula (8.2). With $n_1 = n_2 = n$, we will have $t_1 = t_2 = t'$ and formula (8.2) will give $t_{.05} = t'$, where t' is the tabled value of t with $n_1 - 1 = n_2 - 1 = n - 1$ d.f. Thus, with equal n's and heterogeneity of variance, we may calculate t in the usual way, but the obtained value of t should be evaluated in terms of the tabled value for ½ the number of degrees of freedom we would have with homogeneity of variance.

CONDITIONS MAKING FOR HETEROGENEITY OF VARIANCE

Under what conditions may we expect a given treatment to influence the variance of the observations? To consider this question, suppose that for a control or standard condition, in the absence of any treatment effect, the variance of the measures of the dependent variable for a sample of n observations is s_1^2. Let any given value of X, for the control condition, be designated by X_1. Now suppose we obtain the same number of observations for a given treatment or experimental condition. Let any given value of X, for the treatment, be designated by X_2, and the variance of the obser-

vations, for the treatment, be designated by s_2^2. By means of the F test, assume we find that the difference between s_1^2 and s_2^2 is statistically significant.

Nonrandom Assignment of Subjects

One possible explanation of the significant difference in the variances is that we did *not* randomly assign the subjects to the two groups. If one of the two groups initially included those subjects more homogeneous in their performance than the other, then we might also expect the two groups to differ in variability at the conclusion of the experiment. We hope to rule out this possible explanation by randomly assigning the subjects to the two groups. Thus, if the two variances differ significantly, we may be justified in assuming that the difference is the result of the treatments. We say we *may* be justified in this assumption because, even with random assignment, we know that we will occasionally obtain a significant difference between the two variances simply as a result of chance.

Nonadditivity

If the variances for two treatment groups differ significantly, this may be because the treatment effects are not additive. By *additive* we mean that if X_1 is the value of a given observation under a control condition, then under the treatment condition we would have

$$X_2 = X_1 + a$$

where a represents a constant treatment effect. If a treatment effect is additive, then it can easily be shown that

$$s_2^2 = s_1^2$$

since the addition or subtraction of a constant has no influence upon the variance. On the other hand, suppose that the treatment, instead of acting in an additive fashion, acts in a multiplicative fashion. Then we would have

$$X_2 = X_1 a$$

and it can easily be shown that the variance of X_2, in relation to the variance of X_1, will be

$$s_2^2 = s_1^2 a^2$$

Thus, if the treatment does not operate additively, but rather multiplicatively, then we may expect to find that s_1^2 and s_2^2 will differ.[3]

[3] If treatment effects are nonadditive, then a transformation of the original scale of measurement may provide a new scale on which the treatments are additive. Various transformations are discussed in the next chapter.

Treatments Operating Differentially with Respect to Organismic Variables

Let us consider another possible explanation for a significant difference between s_1^2 and s_2^2. Suppose we find that s_2^2 is significantly greater than s_1^2. A possible explanation of this result is that the treatment operates *differentially* with respect to an organismic variable. For example, if we have divided a group of n subjects, at random, into two groups of n_1 and n_2 subjects each, we would expect that these two groups, if tested under identical conditions, would not differ significantly in their variances. Furthermore, if we were to obtain a measure of anxiety, perhaps by means of the Taylor Manifest Anxiety Scale (MAS), prior to the experiment itself, we would expect the two groups to show only chance or random differences in their mean scores on the MAS. This should be true, if we have randomly assigned the subjects to the two groups.

Suppose, however, that the treatment operates differentially upon those subjects with high anxiety scores and upon those subjects with low anxiety scores. To be specific, let us assume that X is a measure of performance on a pursuit meter and that the treatment of interest is this performance under a condition of stress. For the stress condition, let us assume that each subject is given an electric shock each time the stylus makes a contact. Now, suppose high and low anxiety subjects react differentially to shock. Assume, for example, high anxious subjects go to pieces and thus have a considerably lower performance under the treatment than they would if tested under a control or normal condition For these subjects, we would have $X_2 = X_1 - a$, where a represents a constant treatment effect subtracting from performance.

Let us also assume that subjects with a low degree of anxiety may be of such a nature that they do their best under conditions of stress or shock. For these subjects let a be a constant treatment effect adding to performance. Then for the low anxious subjects, we would have $X_2 = X_1 + a$, that is, their performance under the experimental condition would, in general, be improved.

Thus, with the assumptions we have made and assuming also that degree of anxiety does not influence performance in the control group, we would have the distribution of X_2 measures for the treatment group extending in both directions from the mean over the corresponding measures for the control group. The high anxious subjects, in the treatment group, would, for example, be expected to have a much lower mean and the low anxious subjects a much higher mean than the mean of the control group. Since the subjects in the treatment group are being moved, differentially, in both directions from the control group mean, we would expect the treatment group to have a greater variance than the control group.

It is difficult to overemphasize the importance of organismic variables in accounting for differences in variability of performance of subjects under different experimental conditions. With random assignment, and if the treatment effects are not multiplicative, it seems that one of the most probable explanations for significant differences in variances is that of the differential operation of a given treatment upon differences in an organismic variable. In many psychological experiments, it may be of considerable value to obtain measures of one or more relevant organismic variables. To find that subjects with different values of an organismic variable react differentially to a given treatment is of perhaps even more psychological importance than to find that all subjects respond to the treatment in the same manner.

ASSUMPTIONS OF THE *t* TEST

Normality of the Population Distribution

In our discussion of the *t* test for the difference between two means, we have assumed that the dependent variable X is normally distributed in the population. If X is not normally distributed, will we be in error in the conclusions drawn by using the *t* test to evaluate the difference between the two treatment means? There is considerable evidence to indicate that the *t* test is relatively insensitive to *certain* departures from normality in the population sampled. We have already seen, in the case of the normal curve or *z* test, that the important consideration is not the nature of the distribution in the population, but rather the nature of the random sampling distribution of the statistic of interest. The population distribution is of importance only insofar as it may introduce nonnormality in the sampling distribution.

In experimental work, our interest is primarily in the treatment means and the difference between the means. More specifically, in tests of significance, we are interested in the probabilities associated with the random sampling distribution of the mean or the difference between two means. If the population distribution is normal, the random sampling distribution of the mean is also normal, as is also the difference between two means when both populations are normal. But, fortunately, a normally distributed population is not the only population for which the random sampling distribution of the mean is normal or approximately normal. According to a very important theorem, based upon probability theory and called the *central limit theorem*, the random sampling distribution of the mean of n observations drawn from *any* population with population mean m and variance σ^2 approaches the normal distribution as n becomes large. How large n must be before the sampling distribution of the mean is sufficiently normal so that the probabilities associated with the sampling distribution

can be approximated by the t test depends upon the form of the population distribution.

We have already seen that for a rectangular population such as the binomial with $P = Q$, that $n = 10$ is sufficiently large to approximate the probabilities associated with the sampling distribution of p, the mean of a sample from a binomial population. If the binomial population is skewed with $P \neq Q$, then, as a general rule we have said that we want both nP and nQ to be equal to or greater than 5, before using a test of significance which assumes p to be normally distributed. Thus, obviously, in this instance, the greater the degree of skewness, the larger we want n, before using a test of significance based upon the assumption of normality.

If the population is rectangular so that we have a greater range of values of X rather than just two values as in the binomial population, the probabilities associated with the sampling distribution of the mean can be approximated quite well, even with n less than 10, by assuming the sampling distribution to be normal. For example, if the population is distributed rectangularly over the values 1, 2, 3, 4, and 5, the probabilities associated with the sampling distribution of the mean, when $n = 2$, can be approximated fairly well by assuming the means to be normally distributed. The approximation will be better as n increases. Similar considerations are applicable if the population distribution is skewed. In general, for moderate degrees of skewness, the sampling distribution of the mean will be approximately normal for relatively small samples and larger samples will be required as the degree of skewness increases.

In experimental work, if the population of interest is not normal, the departures from normality will often be of the kinds described. We will have either a skewed distribution or the distribution may be somewhat flatter than the normal distribution. Under these circumstances, we shall not often be in error in tests involving means by using the t test, if the number of observations in each sample is sufficiently large. The probabilities associated with the t test, in this instance, may not be exactly equal to those that are tabled and which were derived by assuming sampling from a normal population, but the exact probability of a difference between two means is pertinent only if the test of significance gives a borderline result. Assume, for example, that we have chosen $\alpha = .05$ and we have a highly significant result, with P much smaller than .05. Then the conclusion that we make on the basis of the test of significance is not likely to be changed, even if we did know the exact probability associated with the difference. The true probability may be somewhat larger or somewhat smaller, yet it is not likely to be sufficiently greater than our approximation of it to cause us to change the inference made on the basis of the approximation. It is unlikely, in other words, that the true probability will exceed

.05 and thus result in a reversal of our decision about the significance of our result.

Continuity of the Dependent Variable

One further point needs to be considered with respect to X, the dependent variable. As we have indicated, we assume X to be *continuously* distributed in the population. If X is a continuous variable, then the sampling distribution of the mean will also be continuous. That this assumption is not crucial for our tests of significance, we have already seen in the case of the binomial population which is discrete. We corrected for the discreteness of the values of the variable, in this instance, by introducing a correction for discontinuity or a correction for continuity, as it is also called. As n, the number of observations in the sample, becomes large, the correction becomes of less importance, since, as n increases, the discreteness of the sample values of p decreases. Thus with $P = .5$ and $n = 10$, the possible values of p are .00, .10, .20, $\cdots$, 1.00. With $n = 20$, the possible values of p are .00, .05, .10, .15, $\cdots$, 1.00. As n increases, the gaps between the possible values of p become smaller and smaller, and a correction for discontinuity is obviously of less importance for large n's than for small n's Similar considerations apply to the apparent discreteness of psychological test scores and other measures which are obtained in experimental work.

Nonparametric Tests

The discussion above is relevant to problems of evaluating experimental data since new developments in statistical analysis have, within recent years, resulted in a variety of tests of significance generally referred to as *nonparametric* tests.[4] It is characteristic of some of the nonparametric tests, proposed as substitutes for the t test, that they do not involve the assumption of sampling from a continuous and normally distributed population. It has been claimed, therefore, by some advocates of nonparametric tests, that the nonparametric tests are much more appropriate for the analysis of results of psychological experiments than, for example, the t test. The reasons usually cited are that psychological measures are often not continuous and that, in general, they are often not normally distributed. We have argued, however, that these so-called conditions for the use of the t test, and other tests to be discussed later, are not crucial What is crucial is the central limit theorem which concerns the sampling distribution of the mean.

Nonparametric tests are useful and valuable additions to methods of data analysis. Many psychologists and other research workers have become

[4] Nonparametric tests are also referred to as *distribution-free* tests.

aware of the existence of these tests within recent years That the nonparametric tests are, in many respects, "new" additions to the more familiar forms of data analysis, may perhaps account for some of the misunderstanding of their nature and limitations. There is evidence in the experimental literature that nonparametric tests are regarded by some research workers as the only appropriate tests; they seem to believe that a nonparametric test should always be used in preference to a t test. What does not seem to be so well known is that the use of a nonparametric test, when a t test is appropriate, results in a test of significance that has considerably less power than the t test. That is, the inappropriate use of a nonparametric test when the t test is appropriate may result in a failure to detect a difference between two means that would be detected by the t test.

QUESTIONS AND PROBLEMS

1. We have the following measures obtained for a control and for an experimental group:

Control Group		Experimental Group			
11	15	4	15	10	10
11	10	4	3	7	12
10	8	8	13	6	14
12	10	9	9	1	8
8	8	12	9	5	5

(*a*) Can we assume that the two variances do not differ significantly? (*b*) Evaluate the significance of the difference between the means.

2. In another investigation, we have the following measures obtained for a control and experimental group:

Control Group				Experimental Group			
12	10	16	11	15	4	10	15
8	13	10	10	12	4	7	3
12	11	11	9	10	8	6	13
11	9	10	9	5	9	1	9
11	11	9	7	8	12	5	9

(*a*) Can we assume that the two variances do not differ significantly? (*b*) Evaluate the significance of the difference between the means.

3. Twenty-five rats were deprived of food for a period of 22 hours before a test trial, but were permitted to satisfy thirst immediately before the test trial. Another group of 25 rats was deprived of food for 22 hours and was permitted to satisfy thirst 12 hours before the test trial. The mean number of responses on the test trial for the first group was 21.36, and the variance was equal to 147.57. In the second group, the mean was 32.92, and the variance was equal to 489.91. The data are from Kendler (1945). (*a*) Can we assume that the two variances

do not differ significantly? (*b*) Evaluate the significance of the difference between the means.

4. A group of 47 rats was tested in a maze placed in a room with temperature at 55 to 58 degrees Fahrenheit. Another group of 46 rats was tested in a room with temperature at 75 to 79 degrees Fahrenheit. The mean number of trials required to learn a maze in the "cold" room was 19.8 with s equal to 7.36. The mean number of trials required for learning in the "normal" room was 25.9 with s equal to 13.3. The data are from Moore (1944). (*a*) Can we assume that the two variances do not differ significantly? (*b*) Evaluate the significance of the difference between the means.

5. A group of 78 subjects was taught shorthand by the "word" method, and another group of 108 subjects was taught by the "sentence" method. At the end of the first semester, both groups were tested on a word test consisting of a list of words dictated slowly by the instructor and written in shorthand by the students. The mean score for the "word" group was 31.72 with s equal to 8.01. The mean score for the "sentence" group was 35.52 with s equal to 22.93. The data are from Clark and Worcester (1932). (*a*) Can we assume that the two variances do not differ significantly? (*b*) Evaluate the significance of the difference between the means. (*c*) If subjects were not randomly assigned to the two treatments, what bearing would this have upon the interpretation of the results?

6. Scores on a visual-motor test were obtained for a group of 70 "control" psychiatric cases with diagnoses other than cerebral brain damage. Another group of 70 psychiatric cases with diagnoses of cerebral brain damage was also tested. The mean score for the "control" group was 3.5 with s equal to 4.8. The mean score for the "brain damage" group was 11.6 with s equal to 7.3. The data are from Graham and Kendall (1946). (*a*) Determine whether the variances and the means for the two groups differ significantly. (*b*) In the absence of randomization in the assignment of subjects to the two groups, how would you interpret the results of the tests of significance?

7. In a study by French and Thomas (1958) a group of 92 subjects was divided into two groups on the basis of their scores on an achievement test. The 47 subjects with scores of 8 or higher on the test are called the "high" group and the 45 subjects with scores of 7 or lower are called the "low" group. Both groups were given a problem to solve and one of the variables measured was the time spent on the task. For the "high" group the mean was 27.14 minutes with a standard deviation of 10.02 minutes. For the "low" group the mean was 13.79 minutes with a standard deviation of 8.13 minutes. (*a*) Determine whether the variances and the means for the two groups differ significantly. (*b*) In the absence of randomization in the assignment of subjects to the two groups, how would you interpret the results of the tests of significance? (*c*) Is it possible to conclude that the "level" of achievement produced the difference in the mean times? (*d*) What are some other possible organismic differences between the two groups?

8. We have randomly assigned subjects to two groups so that we have $n = 20$ in each group. (*a*) What value of t will be required for significance at the 5 per cent level if the variances differ significantly? (*b*) What value of t will be required for significance at the 5 per cent level if the variances do not differ significantly?

9. For each of the above problems in which you found a significant difference between the two variances, discuss possible conditions which may account for the heterogeneity of variance.

10. Examine a recent issue of a journal which publishes the results of research. Try to find an article in which the investigator reports a significant difference in the variances of his two groups or in which you can demonstrate that the two variances differ significantly. If he has not used the tests described in this chapter, reanalyze his data and see whether any conclusions would be changed. Discuss possible conditions which may account for the heterogeneity of variance.

11. Examine a recent issue of the *Journal of Experimental Psychology*. Consider only those studies which can be considered "comparative," that is, in which the major objective is to compare a difference between treatments. If you find a case where randomization was not used in a comparative experiment, what interpretation does the investigator place upon his findings? Would you agree with his interpretation?

12. Define, briefly, each of the following terms:

additivity of treatment effects	nonparametric test
central limit theorem	homogeneity of variance

9

INTRODUCTION TO THE ANALYSIS OF VARIANCE

INTRODUCTION

In the last two chapters we have discussed the application of the t test to problems involving the significance of the difference between the means of two independent samples. We considered the null hypothesis $m_1 = m_2$ under two conditions. In the first instance, we discussed procedures for testing the null hypothesis $m_1 = m_2$ when the samples offered no significant evidence against the null hypothesis $\sigma_1^2 = \sigma_2^2$. In the second instance, we discussed procedures for testing the null hypothesis $m_1 = m_2$ when the samples did offer significant evidence against the null hypothesis $\sigma_1^2 = \sigma_2^2$, that is, when the latter hypothesis was rejected. We are now ready to consider methods that can be used to test the significance of the differences between three or more means. The technique we shall use is known as the *analysis of variance.*

The early development of the analysis of variance as a powerful tool in experimental and research work was largely the accomplishment of Sir R. A. Fisher and his associates in England. In commenting upon a paper presented by Wishart (1934) before the Royal Statistical Society, Fisher (1934, p. 52) had this to say concerning the analysis of variance:

> We were together learning how to use the analysis of variance, and perhaps it is worth while stating an impression that I have formed—that the analysis of variance, which may perhaps be called a statistical method, because that term is a very ambiguous one—is not a mathematical theorem, but rather a convenient method of arranging the arithmetic. Just as in arithmetical text-books—if we can recall their contents—we were given rules for arranging how to find the greatest common measure, and how to work out a sum in practice, and were drilled in the arrangement and order in which we were to put the figures down, so with the analysis of variance; its one claim to attention lies in its convenience. It is convenient in two ways: (1) because it brings to the eyes and to the mind a summary of a mass of statistical data in which the logical content of the whole is readily appreciated. Probably everyone who has used it has found that com-

parisons which they have not previously thought of may obtrude themselves, because there they are, necessary items in the analysis. (2) Apart from aiding the logical process, it is convenient in facilitating and reducing to a common form all the tests of significance which we may want to apply. I do insist that its claim to attention rests essentially on its convenience. Nearly always we can, if we choose, put our data in other forms and other language. Naturally, like other logical arrangements, it is based on mathematical theorems previously proved, and in particular the tests of significance were based on problems of distribution the solution of which was published for the most part from 1921 to 1924.

That the analysis of variance has proved to be not only a convenient method, as Fisher says, but also a powerful method of analysis for the research worker is demonstrated by the extent to which it has been and is being used in the planning, design, and analysis of research in a variety of disciplines.

CALCULATIONS FOR A RANDOMIZED GROUPS DESIGN

We shall illustrate the necessary calculations in the analysis of variance for a *randomized groups design* in which $n = 40$ subjects have been assigned at random to one of $k = 5$ treatments with 8 subjects for each treatment.[1] The measures on the dependent variable for the 5 groups of subjects are given in Table 9.1.

Table 9.1 Randomized Groups Design with 5 Treatments and 8 Subjects Randomly Assigned to Each Treatment

	Treatments					
Observations	1	2	3	4	5	
1	16	16	2	5	7	
2	18	7	10	8	11	
3	5	10	9	8	12	
4	12	4	13	11	9	
5	11	7	11	1	14	
6	12	23	9	9	19	
7	23	12	13	5	16	
8	19	13	9	9	24	
$\sum X$	116	92	76	56	112	452
$\sum X^2$	1,904	1,312	806	462	1,784	6,268

[1] It is not necessary in the application of the analysis of variance that we have equal n's in each group in a randomized groups design. The presentation, however, is simplified by having equal n's. Furthermore, if each of the treatments is considered to be of equal importance, it is advantageous to have equal n's in the various treatment groups.

Total Sum of Squares

We first determine the *total sum of squares* for the 40 observations, ignoring the fact they have been classified according to the particular treatments. This sum of squares will be given by

$$\sum x_t^2 = \sum X^2 - \frac{(\sum X)^2}{n} \tag{9.1}$$

where $n = n_1 + n_2 + \cdots + n_k$, or the total number of observations. For the data of Table 9.1 we have

$$\begin{aligned}\sum x_t^2 &= (16)^2 + (18)^2 + (5)^2 + \cdots + (24)^2 - \frac{(452)^2}{40} \\ &= 6{,}268 - \frac{(452)^2}{40} \\ &= 1{,}160.4\end{aligned}$$

Between-Groups Sum of Squares

We then find a sum of squares which we shall call the *sum of squares between groups* or the *treatment sum of squares.* In general, if we have k groups of observations with $n_1, n_2, \cdots, n_k$ observations in the respective groups, then the sum of squares between groups will be given by

$$\sum x_b^2 = \frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \cdots + \frac{(\sum X_k)^2}{n_k} - \frac{(\sum X)^2}{n} \tag{9.2}$$

For the data of Table 9.1, we have

$$\sum x_b^2 = \frac{(116)^2}{8} + \frac{(92)^2}{8} + \cdots + \frac{(112)^2}{8} - \frac{(452)^2}{40} = 314.4$$

Within-Groups Sum of Squares

If we subtract the sum of squares between groups from the total sum of squares, we obtain a sum of squares which we shall call the *sum of squares within groups,* or *within treatments,* or the *error sum of squares.* Thus

$$\sum x_w^2 = \sum x_t^2 - \sum x_b^2 \tag{9.3}$$

The sum of squares within treatments is a pooled sum of squares based on the variation of the measures within each treatment group about their

respective treatment means. For example, if we consider each of the k groups of Table 9.1 separately, we would have

$$\sum x_1^2 = 1{,}904 - \frac{(116)^2}{8} = 222.0$$

$$\sum x_2^2 = 1{,}312 - \frac{(92)^2}{8} = 254.0$$

$$\sum x_3^2 = 806 - \frac{(76)^2}{8} = 84.0$$

$$\sum x_4^2 = 462 - \frac{(56)^2}{8} = 70.0$$

$$\sum x_5^2 = 1{,}784 - \frac{(112)^2}{8} = 216.0$$

and the sum of these sums of squares is 846.0 and is equal to the sum of squares obtained by subtraction in formula (9.3). Thus we also have

$$\sum x_w^2 = 1{,}160.4 - 314.4 = 846.0$$

Degrees of Freedom and Mean Squares

The results of these calculations are given in Table 9.2. Each of the sums of squares we have calculated has associated with it a specified number of degrees of freedom. For the total sum of squares, we have $n - 1 = 40 - 1 = 39$ d.f. For the sum of squares within groups, we have

Table 9.2 Analysis of Variance for the Randomized Groups Design—Original Data in Table 9.1

Source of Variation	Sum of Squares	d f.	Mean Square	F
Between groups	314.4	4	78.60	3.25
Within groups	846.0	35	24.17	
Total	1,160 4	39		

$n - k = 40 - 5 = 35$ d.f. The degrees of freedom for the sum of squares within groups are based on the following consideration: each of the separate sums of squares, $\sum x_1^2$, $\sum x_2^2$, $\cdots$, $\sum x_k^2$, has $n_i - 1 = 8 - 1 = 7$ d.f. Then, since these k sums of squares have been pooled to obtain the sum of squares within groups, the latter will have $k(n_i - 1) = n - k$ d.f. For the sum of squares between groups, we have $k - 1 = 5 - 1 = 4$ d.f. If we divide the sum of squares within groups and the sum of squares between groups by their respective degrees of freedom, we obtain the two variance estimates, usually called *mean squares*, shown in column (4) of Table 9.2.

Test of Significance

For the randomized groups design, we define F as

$$F = \frac{\text{Mean square between groups}}{\text{Mean square within groups}} \tag{9.4}$$

and this F will have $k - 1$ d.f. for the numerator and $n - k$ d.f. for the denominator. For the data of Table 9.2 we have

$$F = \frac{78.60}{24.17} = 3.25$$

We do not have a row entry corresponding to 35 d.f. in the table of F, but with $\alpha = .05$ we find that the critical value for 4 and 34 d.f. is $F = 2.65$ and for 4 and 36 d.f. the critical value is $F = 2.63$. Thus, it is obvious that our obtained value of 3.25 is significant with $\alpha = .05$ and we reject the null hypothesis $m_1 = m_2 = m_3 = m_4 = m_5$. The differences between the five sample means are sufficiently great that we do not believe they are all estimates of a common population mean.

NATURE OF THE SUMS OF SQUARES IN A RANDOMIZED GROUPS DESIGN

In Table 9.3 we introduce a notation for the observations of Table 9.1. We shall find this notation very convenient in the analysis of variance. Each observation in the table is identified by two subscripts, the first

Table 9.3 Identification of Observations for a Randomized Groups Design with $k = 5$ Treatments and $n = 8$ Observations for Each Treatment

	Observations								
Treatments	1	2	3	4	5	6	7	8	Means
1	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	$\bar{X}_{1.}$
2	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{26}	X_{27}	X_{28}	$\bar{X}_{2.}$
3	X_{31}	X_{32}	X_{33}	X_{34}	X_{35}	X_{36}	X_{37}	X_{38}	$\bar{X}_{3.}$
4	X_{41}	X_{42}	X_{43}	X_{44}	X_{45}	X_{46}	X_{47}	X_{48}	$\bar{X}_{4.}$
5	X_{51}	X_{52}	X_{53}	X_{54}	X_{55}	X_{56}	X_{57}	X_{58}	$\bar{X}_{5.}$

corresponding to a particular treatment and the second to a particular observation for the treatment. Thus X_{32} is the second observation for Treatment 3. We let X_{kn} be a general symbol for any observation, with the understanding that k and n when used as subscripts may correspond to variables. For the data of Table 9.3, k can take any value from 1 to 5, since there are 5 treatments, and n can take any value from 1 to 8, since

there are 8 observations for each treatment. When k and n are used alone or as coefficients of other terms, they will always represent constants.

The over-all mean of the $kn = 40$ observations will be represented by $\bar{X}..$, where the dots indicate that we have summed over all values of X_{kn}. The various treatment means can be represented by $\bar{X}_{1.}$, $\bar{X}_{2.}$, $\bar{X}_{3.}$, $\bar{X}_{4.}$, and $\bar{X}_{5.}$, where the dot which has replaced the subscript n means that we have summed over the n observations for a given treatment. Then $\bar{X}_{k.}$ will be a general symbol for any given treatment mean. Since k is a subscript, it can, for the present example, take any value from 1 to 5. The deviation of a given mean from the over-all mean will be represented by $a_{k.} = \bar{X}_{k.} - \bar{X}..$ and, since k is a subscript, $a_{k.}$ is a variable. In the present example, we would have five such deviations. We also define the deviation of a given value from the over-all mean as $x = X_{kn} - \bar{X}..$ and the deviation of a given value from the mean of the group to which it belongs as $x_k = X_{kn} - \bar{X}_{k.}$. In summary, then

k = the number of treatments
n = the number of observations for each treatment
kn = the total number of observations
X_{kn} = an observation on the dependent variable
$\bar{X}..$ = the over-all mean of all kn observations
$\bar{X}_{k.}$ = the mean of any given treatment group of n observations
$a_{k.} = \bar{X}_{k.} - \bar{X}..$
$x = X_{kn} - \bar{X}..$
$x_k = X_{kn} - \bar{X}_{k.}$

Then, considering a single sample or treatment group, we have

$$x = X_{kn} - \bar{X}..$$

$$x - x_k = (X_{kn} - \bar{X}..) - (X_{kn} - \bar{X}_{k.})$$

$$x = x_k + a_{k.}$$

$$x^2 = {x_k}^2 + 2x_k a_{k.} + {a_{k.}}^2$$

Summing over the n observations in the sample, we have

$$\sum_1^n x^2 = \sum_1^n {x_k}^2 + 2a_{k.} \sum_1^n x_k + \sum_1^n {a_{k.}}^2$$

But, since $\sum_1^n x_k = \sum_1^n (X_{kn} - \bar{X}_{k.}) = 0$, and since ${a_{k.}}^2$ is summed n times, once for each of the n observations in the sample, we have

$$\sum_1^n x^2 = \sum_1^n {x_k}^2 + n{a_{k.}}^2$$

For each of the k samples we shall have an expression similar to the one above. Summing over all k samples, we have

$$\sum_1^k \sum_1^n x^2 = \sum_1^k \sum_1^n x_k{}^2 + n \sum_1^k a_k{}^2 \tag{9.5}$$

The first term on the left in (9.5) above will be equal to $\sum_1^k \sum_1^n (X_{kn} - \bar{X}..)^2$ and is what we have called the *total sum of squares.* The total sum of squares measures the variation of the observations about the over-all mean The first term on the right is equal to $\sum_1^k \sum_1^n (X_{kn} - \bar{X}_{k.})^2$ and is what we have called the *sum of squares within groups.* The sum of squares within groups is a pooled sum of squares based upon the variation of the n measures in each group about the mean of the group to which they belong. The last term on the right is equal to $n \sum_1^k (\bar{X}_{k.} - \bar{X}..)^2$ and is a weighted sum of squares based upon the variation of the group means about the over-all mean.

What we have shown is that whenever we have k groups of n observations each, it is always possible to analyze the total sum of squares into two parts: the sum of squares within groups and the sum of squares between groups. The degrees of freedom associated with the total sum of squares will be $kn - 1$. The sum of squares within groups will have $kn - k = k(n - 1)$ d.f. and the sum of squares between groups will have $k - 1$ d.f.

MEAN SQUARES AND THE TEST OF SIGNIFICANCE IN A RANDOMIZED GROUPS DESIGN

Mean Square Within Groups

If the k samples are drawn at random from normal populations with identical variances so that each $\sigma_k{}^2 = \sigma^2$, then each of the samples will provide a separate estimate $s_k{}^2$ of the same population variance σ^2. Combining these estimates by means of formula (7.13) gives us a single estimate. Thus

$$s^2 = \frac{\sum x_1{}^2 + \sum x_2{}^2 + \cdots + \sum x_k{}^2}{k(n-1)} = \frac{\sum_1^k \sum_1^n x_k{}^2}{k(n-1)}$$

The numerator of the above expression is the sum of squares within groups and the denominator is the number of degrees of freedom associated with the sum of squares within groups. Thus s^2 is identical with the mean square within groups.

Mean Square Between Groups

Consider now the mean square between groups. If the k samples are drawn at random from the same normal population, or from normal populations with identical means, so that each $m_k = m$, then each of the k sample means will provide a separate estimate of the same population mean m. Then we can combine the observations to obtain a single estimate

$$\bar{X}_{..} = \frac{n_1\bar{X}_{1.} + n_2\bar{X}_{2.} + \cdots + n_k\bar{X}_{k.}}{n_1 + n_2 + \cdots + n_k}$$

or

$$\bar{X}_{..} = \frac{\sum X_{1.} + \sum X_{2.} + \cdots + \sum X_{k.}}{kn} \tag{9.6}$$

and the value obtained from formula (9.6) will be the over-all mean of the kn observations. Then

$$s_{\bar{x}}^2 = \frac{(\bar{X}_{1.} - \bar{X}_{..})^2 + (\bar{X}_{2} - \bar{X}_{..})^2 + \cdots + (\bar{X}_{k.} - \bar{X}_{..})^2}{k - 1}$$

or

$$s_{\bar{x}}^2 = \frac{\sum_1^k a_{k.}^2}{k - 1} \tag{9.7}$$

will be an estimate of $\sigma_{\bar{x}}^2$.

We have already seen that the variance of the means of random samples of n observations each drawn from the same population is estimated by

$$s_{\bar{x}}^2 = \frac{s^2}{n}$$

and thus

$$ns_{\bar{x}}^2 = s^2 \tag{9.8}$$

If we multiply both sides of formula (9.7) by n, the number of observations in each sample, we have

$$ns_{\bar{x}}^2 = \frac{n[(\bar{X}_{1.} - \bar{X}_{..})^2 + (\bar{X}_{2.} - \bar{X}_{..})^2 + \cdots + (\bar{X}_{k.} - \bar{X}_{..})^2]}{k - 1}$$

or

$$ns_{\bar{x}}^2 = \frac{n\sum_1^k a_{k.}^2}{k - 1} \tag{9.9}$$

and the right-hand side of formula (9.9) is also an estimate of the common population variance σ^2.

The numerator of formula (9.9) is identical with the sum of squares between groups and the denominator is the number of degrees of freedom associated with the sum of squares between groups. Thus, formula (9.9) is identical with the mean square between groups in the analysis of variance.

Test of Significance

When we find F, by dividing the mean square between groups by the mean square within groups, we have a ratio between two variance estimates. If the null hypothesis is true, then the numerator of the F ratio should exceed the denominator only as a result of random sampling. With homogeneity of the sample variances, we will obtain a significant value of F if the sample means vary more than to be expected in random sampling from populations with identical means. If we draw samples at random from the same population or from populations with identical means, then the sample means should vary only within the limits of random sampling. On the other hand, if the samples are drawn from populations in which the means are not identical, this will serve to increase the variation in the means, as measured by formula (9.7). Thus, the mean square between groups will tend to be larger than the mean square within groups. A significant value of F, as given by formula (9.4) is taken as evidence that the population means are not equal.

To know merely that the k means in the set differ significantly is alone not very satisfying. We often wish to know something more specific about the nature of the differences. In a subsequent chapter, we shall show additional tests that may be applied to the set of k means.

HETEROGENEITY OF VARIANCE

Our earlier discussion of heterogeneity of variance, in connection with the t test, is pertinent also to the analysis of variance. To obtain the mean square within groups, we combined the separate variance estimates of the k samples under the assumption that they were all estimates of the same population variance. In some experiments it may appear that the separate sample variances are quite dissimilar and we may wish to determine whether the null hypothesis $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$ is tenable before proceeding with the analysis of variance and test of significance concerning the means.

Bartlett's Test with Equal d.f.'s

A test for homogeneity of k variances has been described by Bartlett (1937) and we show the necessary calculations for the problem we have already treated by the analysis of variance. The original data are given in

Table 9.1. The sums of squares for the five treatment groups are given in Table 9.4. In the fourth column of the table we have entered the variance

Table 9.4 Bartlett's Test of Homogeneity of Variance for k Variances with Equal Degrees of Freedom

Treatment	d f	$\sum x_k^2$	s_k^2	$\log s_k^2$
1	7	222.0	31.71	1.50120
2	7	254.0	36.29	1 55979
3	7	84.0	12.00	1.07918
4	7	70.0	10 00	1.00000
5	7	216.0	30.86	1.48940
$\sum$			120 86	6 62957

Computations:

1. $\dfrac{\sum s_k^2}{k} = \dfrac{120.86}{5} = 24.17$; $\log \dfrac{\sum s_k^2}{k} = 1.38328$
2. $k \log \dfrac{\sum s_k^2}{k} = (5)(1.38328) = 6.91640$
3. Diff. $= k \log \dfrac{\sum s_k^2}{k} - \sum \log s_k^2 = 6.91640 - 6.62957 = .28683$
4. $\chi^2 = (2.3026)(n-1)(\text{Diff.}) = (2.3026)(7)(.28683) = 4.623$
5. Correction $= 1 + \dfrac{k+1}{3k(n-1)} = 1 + \dfrac{6}{(3)(5)(7)} = 1.057$
6. Corrected $\chi^2 = \chi^2/\text{Correction} = 4.623/1.057 = 4\,374$

s^2 for each group. The variances are obtained by dividing the sums of squares by the corresponding degrees of freedom. In the last column, the logarithms of the variances have been entered.

If the null hypothesis is true, then the separate values of s^2 should not differ any more than is to be expected in random sampling from a common population with variance σ^2. The test of significance for the null hypothesis is made by means of χ^2, and the number of degrees of freedom for evaluating χ^2 will be $k - 1$, where k is the number of independent variance estimates.

The value of χ^2 calculated in line 4 of Table 9.4 is somewhat biased in that it tends to exaggerate the significance level. If the value of χ^2 as found in line 4 is significant, then the "correction" as shown in line 5 should be calculated. Then the "corrected" χ^2 obtained in line 6 will give a more accurate value for interpretation. The corrected χ^2 will always be less than the value obtained in line 4. Thus, if the value first found is not significant, there is no need to apply the correction. In the present problem, for example, since the χ^2 required for significance, with $\alpha = .05$, for $k - 1 = 4$ d.f. is 9.488, there would really be no necessity of obtaining the corrected value

of χ^2. It is obvious that the obtained value of 4.623 is not significant, and since the corrected value will be even smaller, it also will not be significant.

The χ^2 test, applied to the variances of the samples, gives support to the belief that these samples are not heterogeneous in variance. Since the obtained value of χ^2 is not significant, the data offer no significant evidence against the hypothesis that these samples were drawn from populations with equal variances.

Bartlett's Test with Unequal d.f.'s

The test for homogeneity of variance when the separate estimates have differing degrees of freedom is illustrated in Table 9.5. The value of

Table 9.5 Bartlett's Test for Homogeneity of Variance for k Variances with Unequal Degrees of Freedom

Treatments	d.f.	$\frac{1}{\text{d.f.}}$	$\sum x_k^2$	s_k^2	$\log s_k^2$	(d f.) ($\log s_k^2$)
1	20	.05000	244.0	12.20	1.08636	21.72720
2	12	.08333	162.0	13.50	1.13033	13.56396
3	14	.07143	110.0	7.86	.89542	12.53588
4	9	.11111	98.0	10.89	1.03703	9.33327
$\sum$	55	31587	614.0			57.16031

Computations:

1. $\frac{\sum\sum x_k^2}{\sum \text{d.f.}} = \frac{614}{55} = 11.16$; $\log \frac{\sum\sum x_k^2}{\sum \text{d.f.}} = 1.04766$
2. $\sum \text{d.f.} \left(\log \frac{\sum\sum x_k^2}{\sum \text{d.f.}} \right) = (55)(1.04766) = 57.62130$
3. $\text{Diff.} = \sum \text{d.f.} \left(\log \frac{\sum\sum x_k^2}{\sum \text{d f.}} \right) - \sum \text{d.f.} \ (\log s_k^2)$
   $= 57.62130 - 57.16031 = .46099$
4. $\chi^2 = (2.3026)(\text{Diff.}) = (2.3026)(.46099) = 1.061$
5. $\text{Correction} = 1 + \left[\frac{1}{(3)(k-1)} \right] \left[\sum \frac{1}{\text{d.f.}} - \frac{1}{\sum \text{d.f}} \right]$
   $= 1 + \left[\frac{1}{(3)(3)} \right] \left[.31587 - \frac{1}{55} \right] = 1.033$
6. Corrected $\chi^2 = \chi^2/\text{Correction} = 1.061/1.033 = 1.027$

χ^2, in this instance, will also have $k - 1$ d.f., where k is the number of independent variance estimates. Since it is obvious here also that the value of χ^2 obtained in line 4 is not significant, with $\alpha = .05$, the application of the correction is not necessary. In the case of borderline significance of the value of χ^2 obtained in line 4, the final decision as to significance should be based upon the corrected value obtained in line 6.

Sensitivity of Bartlett's Test to Nonnormality

Caution must be exercised in interpreting a significant value of χ^2 as indicating that it is necessarily the variances that differ. Box (1953), for example, has shown that Bartlett's test is as sensitive to nonnormality as to differences in variances. Thus, the test can be safely interpreted as indicating that it is the variances that differ only if we have assurance that normality is present.[2] The F test for means is, like the t test, remarkably insensitive to nonnormality of the population distribution, provided the departures from normality are of the same kind for the various populations sampled. Thus, for example, if the population of observations represented by one treatment is skewed and if the populations for the various other treatments are skewed in the same direction, the F test will be primarily sensitive to differences in means and not to the skewness.

TRANSFORMATIONS OF SCALE[3]

Square Root Transformations

The presence of correlation between the variances and means of the treatments is one indication of departure from normality, and this is likely to be associated with heterogeneity of variance. For a particular type of distribution, called the Poisson distribution, the mean and variance are equal, that is, $m = \sigma^2$. Thus samples drawn from different Poisson distributions may be expected to differ with respect to both means and variances. Poisson distributions are likely to be obtained when the observations consist of counts, such as the number of responses of some kind made in a fixed period of time. If the distribution of observations is such that the means and variances of the treatments tend to be proportional or correlated, then a transformation of the original observations to a new scale may stabilize the variance. For the Poisson distribution, the transformation recommended by Bartlett (1936) is the square root transformation. For example, Bartlett suggests that we should transform each value of X by taking $\sqrt{X + .5}$. Freeman and Tukey (1950) have suggested that for the Poisson distribution the variance stabilizing properties of the square root transformation are improved by taking $\sqrt{X} + \sqrt{X + 1}$.

[2] By examining the distribution of the residuals, $x_k = X_{kn} - \bar{X}_{k\cdot}$, for all kn observations we can obtain some indication as to whether nonnormality is present in the observations under consideration. The cumulative proportion distribution of the residuals can, for example, be plotted on normal probability paper. If a cumulative proportion distribution of a normally distributed variable is plotted on this paper, the resulting graph will be a straight line.

[3] For a discussion of additional transformations, see Bartlett (1947), Mueller (1949), and Curtiss (1943).

To illustrate the two transformations described above, we consider an experiment by Sleight (1948). Sleight was interested in the legibility of readings of various dial types Five different dial types were investigated: horizontal, open window, round, vertical, and semicircular.

We give in Table 9.6 only the data for the round, vertical, and semicircular types. We shall also assume that different subjects were assigned

Table 9.6 Number of Errors Made by 3 Groups of Subjects in Reading 3 Different Dials

	Round	Vertical	Semicircular
	2	6	4
	2	6	2
	0	10	6
	4	12	4
	3	6	7
$\sum X$	11	40	23
Means	2.2	8.0	4.6
s^2	2.2	8.0	3 8

at random to each of the three treatments. For each treatment, we have calculated the mean and variance and these are also given in the table. It is clear that the means and variances of the original data are proportional and highly correlated and for the first two treatments we have means and variances that are identical. This suggests that the square root transformation is appropriate.

Table 9.7 The $\sqrt{X + .5}$ Transformation for the Data of Table 9.6

	Round	Vertical	Semicircular
	1.58	2.55	2.12
	1.58	2.55	1.58
	.71	3.24	2.55
	2.12	3.54	2.12
	1.87	2.55	2.74
$\sum X$	7.86	14.43	11.11
Means	1.57	2.89	2 22
s^2	.28	.22	.20

The transformation $\sqrt{X + .5}$ is given in Table 9.7. The means and variances for the observations on the transformed scale are given at the bottom of the table. It is perfectly clear here that the means and variances

are no longer proportional and that variances are more homogeneous. The analysis of variance could now be applied to the transformed data.

Table 9.8 The $\sqrt{X} + \sqrt{X+1}$ Transformation for the Data of Table 9.6

	Round	Vertical	Semicircular
	3.15	5.10	4 24
	3.15	5.10	3 15
	1.00	6.48	5.10
	4.24	7.07	4.24
	3 73	5 10	5.47
$\sum X$	15.27	28 85	22.20
Means	3 05	5.77	4.44
s^2	1 53	.89	.81

Table 9 8 shows the results obtained with the transformation $\sqrt{X} + \sqrt{X+1}$. Again it is apparent that the transformation has tended to make the variances more homogeneous.[4]

Logarithmic Transformation

Another case in which heterogeneity of variance may be found is that in which the standard deviations of the various treatment groups tend to be proportional to the treatment means. In this case a transformation to a logarithmic scale is recommended by Bartlett (1947). When values of X equal to zero are present, the transformation may take the form log $(1 + X)$.

In a study of the hoarding behavior of rats, Morgan (1945), for example, found that the logarithm of the number of pellets hoarded resulted in distributions which were more approximately normal and with more homogeneous variances. Similarly, Haggard (1945) has found that a logarithmic transformation is suitable for measures of the galvanic skin response.

Inverse Sine Transformation

A transformation has also been suggested for counts based upon samples drawn from different binomial populations. For example, we may have one binomial population in which P, the probability of a successful response on a given trial, is constant from trial to trial We have a fixed number of trials and the value of X consists of the number of successful

[4] Mosteller and Bush (1954) provide a table of the values of $\sqrt{X} + \sqrt{X+1}$ which facilitates the computations for this transformation.

responses. If the treatments involve samples from different binomial populations, with varying values of P, then we may expect the sample means and variances to be related. The transformation suggested in this instance is the inverse sine or angular transformation. Values of $\sin^{-1}\sqrt{p}$ where p is the percentage or proportion of correct responses in a fixed number of trials have been tabled by Bliss (1937), but this reference is not readily available. Bliss's tables have been reproduced, however, by Snedecor (1956) and Guilford (1954). Values of the transformation are also available in the Fisher and Yates (1948) tables.

Reciprocal Transformation

In studying the influence of varying amounts of incentive upon speed of running in two groups of 7 rats each, Crespi (1942) found a significant difference in the variances for a 1-unit and a 4-unit incentive group. $F = s_1^2/s_2^2$ was 72.4 when time was used as a measure of performance. For the 6 and 6 d.f. available, $F = 72.4$ is highly significant. Because of this, and for other reasons which he discusses in some detail, Crespi transformed his unit of measurement to a new scale. Instead of using X, the time required to run the path, he transformed to the scale $1/X$, the reciprocals of the time measures. With this transformation the variances were stabilized, as indicated by the F ratio of 1.06 for the transformed observations.

A reciprocal transformation such as that used by Crespi may prove to be useful in other psychological studies where time is the dependent variable. For example, the transformation may be useful in word-association or reaction-time studies or in studies of problem solving where the time taken to solve the problem is the dependent variable.

FURTHER COMMENTS ON THE *F* TEST AND HETEROGENEITY OF VARIANCE

It is not intended to give the impression that conclusions based upon the analysis of variance applied to original data in which the treatment variances differ will be changed if the data are transformed to another scale on which they are more homogeneous and the analysis of variance is applied to the transformed data. For example, the errors for all 5 dial types (we gave only the results for 3 of the dial types in Table 9.7) in the experiment by Sleight were transformed by means of $\sqrt{X + .5}$. An analysis of variance of the transformed data resulted in exactly the same conclusion concerning significance of the treatment means as that based on the analysis of the original data.

The analysis of the transformed data did result in $F = 11.92$ for the ratio of the treatment mean square and the error mean square, while analysis of the original data gave $F = 8.87$. With $\alpha = .05$, $F = 5.67$ is

significant for the number of degrees of freedom available. Thus, although no conclusions concerning significance were changed by the analysis of the transformed data, the probability associated with the F of the second analysis is smaller than the probability associated with the F of the original analysis.

There is considerable evidence to indicate that in the common case in experimental work where the number of observations is the same for the various treatments, the F test for the means in the analysis of variance is little influenced by heterogeneity of variance.[5] As Box (1953) has emphasized, since the F test is very insensitive to nonnormality and since with equal n's it is also insensitive to variance inequalities, it would be best to accept the fact that it can be used safely under most conditions. The F test of the analysis of variance, in other words, remains a robust test under a variety of violations of the assumptions on which it is mathematically based.

QUESTIONS AND PROBLEMS

1. Data for three treatment groups are given below:

A	*B*	*C*
27	22	37
45	24	38
44	42	25
31	41	47
38	31	23

Analyze the results using the analysis of variance.

2. In the Morgan (1945) experiment, Problem 7, Chapter 3, t was used to evaluate the difference between the means of the two groups. Analyze the same data, using the analysis of variance. You should find that $t^2 = F$.

3. We have below 6 samples which were drawn at random from a sampling box. Analyze the data using the analysis of variance.

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
9	9	6	6	12	10
12	10	12	9	6	6
7	8	9	12	8	8
14	3	9	7	7	12
5	7	8	5	3	9
8	8	10	5	13	4
7	5	2	8	7	8
7	9	10	9	13	7
8	3	9	6	6	6
3	8	12	3	6	2

[5] See, for example, Box (1954*a*).

4. Subjects were assigned at random to one of three treatment groups. Measures obtained on the dependent variable are given below:

A	*B*	*C*
22	21	32
35	44	23
45	35	22
24	40	41
43	35	44
38	22	32
23	50	18
30	28	22

5. In an experiment involving 5 treatments, the following measures were obtained:

A	*B*	*C*	*D*	*E*
13	7	12	10	13
9	4	11	12	6
8	4	4	9	14
7	1	9	7	12
8	10	5	15	13
6	7	10	14	10
6	5	2	10	8
7	9	8	17	4
6	5	3	14	9
10	8	6	12	11

Analyze the results using the analysis of variance.

6. The analysis of variance should prove to be an extremely useful tool in problems which involve a "test of technique," i.e., where the experimenter is not sure that he can reproduce his results. Such failures may be the result of inability to standardize and thus control the conditions of the experiment. They may also be due to unreliable observers or unreliable measuring devices or other factors. The problem cited here happens to involve the observers, and was carried out under the direction of Loucks. Subjects were assigned at random to one of four graduate assistants—here referred to as Operators A, B, C, and D. The operators did not test an equal number of subjects, but each operator observed his particular subjects perform under supposedly the same set of conditions. Records were kept of a number of different variables. The one reported here concerns but one phase of the study, the errors made in making turns in an airplane trainer. The records for each operator for his particular group of subjects are as follows:

Subjects	Operator A	Operator B	Operator C	Operator D
1	6	5	4	8
2	4	3	7	5
3	3	4	3	6
4	7	3	7	7
5	13	3	4	7
6	9	4	8	9
7	4	0	7	7
8	10	3	4	10
9	8	4	8	4
10	9	4	4	3
11	8	3	11	3
12	5	3	13	
13	5	4	9	
14	10	2		
15	9	5		
16	15	3		
17	10	3		
18	6	1		
19	4	2		
20	5			
21	7			

Analyze the data using the analysis of variance.

7. In the above example, there is a tendency for the means of the various groups to be proportional to the standard deviations. (*a*) Find the means and standard deviations to determine this for yourself. (*b*) Test the variances for homogeneity. (*c*) Transform the data to the logarithmic scale $\log (X + 1)$. Find the means and standard deviations on the transformed scale. Are the means and standard deviations still proportional? (*d*) Analyze the data on the transformed scale, using the analysis of variance.

8. Subjects were randomly assigned to three treatment groups. The outcomes of the experiment are given below:

Treatment 1	Treatment 2	Treatment 3
12	0	4
12	0	4
19	2	0
24	4	8
12	4	6
11	5	4
19	5	5
22	0	0
11	1	9
11	3	7

(*a*) Find the mean and variance for each group. Note that they tend to be proportional. (*b*) Analyze the data using the analysis of variance. (*c*) Transform the data to the scale $\sqrt{X + .5}$. Find the mean and variance for each group on

the transformed scale. Has the variance been stabilized by the transformation? (*d*) Repeat the analysis of variance with the transformed data.

9. What are the possible conditions which may introduce heterogeneity of variance in the observations obtained for different groups?

10. Why is it possible to regard the treatment mean square, when the null hypothesis of interest is true, as an estimate of the common population variance?

11. Define, briefly, each of the following terms:

error sum of squares
mean square
randomized groups design
sum of squares within groups
total sum of squares
transformation of scale
treatment sum of squares

10

MULTIPLE COMPARISONS IN THE ANALYSIS OF VARIANCE

INTRODUCTION

Suppose we have tested a set of k means by the analysis of variance and have concluded that the means differ significantly This alone, as we pointed out in the previous chapter, is not very satisfactory. What we would usually like to know is how the means differ. Is every mean significantly different from every other? Are there significant differences between some of the means and not between others?

A variety of methods have been proposed for investigating the differences existing between a set of k means. These test procedures are useful whenever we are concerned with *multiple comparisons* among the means. In this chapter, we shall consider some *selected* methods for multiple comparisons [1]

In making multiple comparisons among the treatment means, it is not necessary that the treatment mean square of the analysis of variance be significant. In other words, we may have a nonsignificant treatment mean square and still use the methods to be described for making multiple comparisons. This does not mean, however, that we should indiscriminately apply the methods, one after another, with the anticipation that one or another may result in some finding that meets the requirements of statistical significance. Our choice of method should, instead, be guided by questions of experimental interest.

DUNCAN'S NEW MULTIPLE RANGE TEST

The first case we shall consider is a comparative experiment in which k treatments are tested. The summary of the analysis of variance for the

[1] The problem of multiple comparisons is not a simple one and the statisticians who have worked upon the problem are not themselves in complete agreement as to procedures. Further discussions can be found in the references cited in this chapter and in Federer (1955). See also the review by Ryan (1959).

Table 10.1 Analysis of Variance for $k = 8$ Treatments with $n = 4$ Observations for Each Treatment

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Between groups	7,803.16	7	1,114 74	30.96
Within groups	864.00	24	36.00	
Total	8,667.16	31		

experiment is given in Table 10.1. The F of 30.96 with 7 and 24 d.f. is significant ($P < .01$).

The observed means, each based upon 4 observations, are given in Table 10.2. We shall assume that we had no a priori hypotheses as to the

Table 10.2 Duncan's New Multiple Range Test Applied to the Differences Between $k = 8$ Treatment Means—The Analysis of Variance for the Same Experiment Is Given in Table 10.1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	Shortest Significant
Means	24.7	41.7	55.6	56.4	60.1	66.3	70.3	77.0	Ranges
A 24.7		17.0	30.9	31.7	35.4	41.6	45.6	52.3	$R_2 = 11.88$
B 41.7			13.9	14.7	18.4	24.6	28.6	35.3	$R_3 = 12.39$
C 55.6				.8	4.5	10.7	14.7	21.4	$R_4 = 12.72$
D 56.4					3.7	9.9	13.9	20.6	$R_5 = 12.96$
E 60.1						6.2	10 2	16.9	$R_6 = 13.17$
F 66 3							4.0	10 7	$R_7 = 13.32$
G 70.3								6.7	$R_8 = 13.44$
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	

Any two treatment means not underscored by the same line are significantly different.

Any two treatment means underscored by the same line are not significantly different.

differences to be found between the 8 means, and in the table they are simply arranged in order of magnitude and identified by the letters A to H. If we wish to determine which of the differences between these means are significant and which are not, the suggested test procedure is Duncan's (1955) *new multiple range test.* We shall illustrate the multiple range test only for the case where we have the same number of observations in each group or for each mean.[2]

[2] An extension of Duncan's new multiple range test for the case of unequal n's is given by Kramer (1956). See also Duncan (1957).

smallest minus the smallest. Since we have ranked the means in order of magnitude, with A being the smallest and H the largest, the order of testing will involve first finding the differences in column (8) of Table 10.2, then those in column (7) and so on, moving from right to left.

Each difference in Table 10.2 is significant if it exceeds the corresponding shortest significant range. If it does not, then it is not significant. The only exception to this rule is that no difference between two means can be significant if the two means are both contained in a subset of means which has a nonsignificant range. The term "subset" may refer to the complete set of means. Because of the exception noted, it is convenient to group these two means and all of the intervening means by underscoring, as shown at the bottom of Table 10.2. No additional tests are made between the means of a subset underscored in the manner described.

Because $H - A$ is the range of 8 means, the difference must exceed $R_8 = 13.44$, the shortest significant range for 8 means. Because $H - B$ is the range of 7 means, it must exceed $R_7 = 13.32$, the shortest significant range for 7 means, and so on. When we come to $H - F$, we find that the difference, 10.7, does not exceed $R_3 = 12.39$. Therefore, this difference is not significant and no further tests are made between the means of the subset F, G, and H. That F, G, and H do *not* differ significantly is shown by the underscoring of these three treatment means at the bottom of Table 10.2.

In column (7) of Table 10.2, the difference $G - E = 10.2$ is a range of 3 means and does not exceed $R_3 = 12.39$. The underscoring of E, F, and G at the bottom of the table shows that these means do *not* differ significantly. In column (6) we find that $F - A = 41.6$, the range of 6 means, exceeds $R_6 = 13.17$, and $F - B = 24.6$, the range of 5 means, exceeds $R_5 = 12.96$. However, $F - C = 10.7$, the range of 4 means, does not exceed $R_4 = 12.72$. No further tests, therefore, are made between the means of the subset C, D, E, and F, and the fact that they form a subset is shown by the underscoring at the bottom of the table.

The final test we make is $B - A = 17.0$, and since this difference exceeds $R_2 = 11.88$, it is significant. The complete results of the various tests are summarized by the underscoring at the bottom of Table 10.2. Any two means *underscored* by the same line do *not* differ significantly. Any two means *not underscored* by the same line *do* differ significantly.

It should be obvious that it is *not necessary to record the complete table of differences*, as we have done in Table 10.2. For example, once we found that $H - F$ was not significant, we know that $H - G$ and $G - F$ are not to be tested. Similarly, having found that $F - C$ is not significant, no tests are to be made between any of the differences in the subset C, D, E, and F, and there would be no need to find any of these differences.

Protection Levels

Duncan's multiple range test is based upon the concept of *protection levels.* A two mean protection level is given by $1 - \alpha$. Thus, if $\alpha = .01$, then the two mean protection level is $1 - .01 = 99$ per cent. If the two population means are in fact equal, then $\alpha = .01$ is the probability that we will wrongly declare them to be significantly different. Our two mean protection level against this erroneous conclusion is $1 - .01 = 99$ per cent. The k-mean protection level for the multiple range test is given by $(1 - \alpha)^{k-1}$. In the present example, with 8 means and $\alpha = .01$, the protection level is $(1 - .01)^{8-1} = 93$ per cent, which is the minimum probability of finding no erroneous significant differences between the 8 means. We thus have somewhat less protection against erroneous conclusions about significance in comparing the differences between the 8 means than we would have if we were testing the significance of the difference between only two means.

The exponent, $k - 1$, for the protection level, is given by the number of independent comparisons which can be made between a set of k means and is equal to the number of degrees of freedom associated with the treatment mean square in the analysis of variance. If we had chosen $\alpha = .05$, then the protection level, based upon degrees of freedom would be, for a set of 8 means, $(1 - .05)^{8-1} = 70$ per cent.

If $k > 2$ means are being compared, it seems reasonable to expect that we are more likely to have some real differences between the means than would be the case if only $k = 2$ means are compared. Duncan has argued, therefore, that in testing the differences between $k > 2$ means, the test of significance should be more powerful, more likely to detect real differences, than when testing the difference between $k = 2$ means. With the multiple range test, the increased power is obtained by risking a lowered protection level as k increases. In essence, the experimenter who uses Duncan's test in evaluating the differences between $k > 2$ means is likely to make fewer Type II errors and somewhat more Type I errors than he would if the protection level for $k > 2$ was the same as that for $k = 2$.

ORTHOGONAL COMPARISONS OF TREATMENT MEANS

If we have an experiment in which k treatments are involved and the results are treated by the analysis of variance, then we shall have a sum of squares between treatments or between groups with $k - 1$ degrees of freedom. It is always possible, as we shall see later, to analyze the sum of squares for treatments into $k - 1$ component parts, each with 1 d.f., such that the sum of these component parts is equal to the sum of squares for treatments. Each of the component parts may involve a comparison be-

tween two or more of the treatments. We shall consider first the case where we have k means with an equal number of observations for each.

Table 10.3 Analysis of Variance for $k = 4$ Treatments with $n = 10$ Observations for Each Treatment

Source of Variation	Sum of Squares	d f.	Mean Square	F
Between groups	83.50	3	27 83	9.09
Within groups	110.16	36	3.06	
Total	193.66	39		

Table 10.3 gives the summary analysis of variance for an experiment in which 4 treatments were tested with $n = 10$ observations for each treatment. $F = 9.09$, with 3 and 36 d.f., is significant ($P < .01$).

Comparisons of the Means

Table 10.4 shows three comparisons that might be made between the group of $k = 4$ means. The numbers given in each column of the table are

Tabel 10.4 Three Orthogonal Comparisons Between $k = 4$ Means with Notation for the Coefficients at the Right and Values of the Coefficients at the Left

(1) Treatment Means	(2)	(3)	(4) Values of Coefficients	(5)	(6)	(7) Coefficients
	$a_{.1}$	$a_{.2}$	$a_{.3}$	$a_{.1}$	$a_{.2}$	$a_{.3}$
$\bar{X}_{1.}$	1	0	½	a_{11}	a_{12}	a_{13}
$\bar{X}_{2.}$	−1	0	½	a_{21}	a_{22}	a_{23}
$\bar{X}_{3.}$	0	−1	−½	a_{31}	a_{32}	a_{33}
$\bar{X}_{4.}$	0	1	−½	a_{41}	a_{42}	a_{43}

called *coefficients* of the treatment means and we shall use a with appropriate subscripts, as shown at the right of the table, to represent these coefficients. The first subscript refers to the particular treatment mean which is to be multiplied by the coefficient and the second subscript corresponds to the particular comparison.

Multiplying the treatment means by the coefficients in the column headed $a_{.1}$, we obtain the comparison

$$d_1 = \bar{X}_{1.} - \bar{X}_{2.}$$

Multiplying the treatment means by the coefficients in the column headed $a_{.2}$, we obtain the comparison

$$d_2 = \bar{X}_{4.} - \bar{X}_{3.}$$

If we multiply the treatment means by the coefficients in the column headed $a_{.3}$, we find that this gives the comparison

$$d_3 = \tfrac{1}{2}(\bar{X}_{1.} + \bar{X}_{2.}) - \tfrac{1}{2}(\bar{X}_{3.} + \bar{X}_{4.})$$

or the difference between the average of the means for Treatments 1 and 2 and the average of the means for Treatments 3 and 4.

Standard Error of a Comparison

Let $\sum a_{.i}$ be the sum of the coefficients in the ith column. Then, if $\sum a_{.i} = 0$, the standard error of the corresponding weighted difference between the means, that is, the difference obtained by multiplying the means by the coefficients in the column, will be

$$s_{d_i} = \sqrt{s^2\left(\frac{a_{1i}^2}{n_1} + \frac{a_{2i}^2}{n_2} + \cdots + \frac{a_{ki}^2}{n_k}\right)} \tag{10.1}$$

where s^2 is the error mean square of the analysis of variance. If the number of observations is the same for each mean, then formula (10.1) may be written

$$s_{d_i} = \sqrt{\frac{s^2}{n}\sum a_{.i}^2} \tag{10.2}$$

where n is the number of observations for a single mean.

Significance of a Comparison

The significance of the difference of the comparison represented by d_i can then be evaluated by finding

$$t = \frac{d_i}{s_{d_i}} \tag{10.3}$$

with d.f. equal to the number of degrees of freedom of the error mean square of the analysis of variance. Confidence limits for d_i may be established in the usual way.

In Table 10.5, we give the means for the analysis of variance reported in Table 10.3, and the coefficients for the three comparisons of Table 10.4. Multiplying the means by the corresponding coefficients in the three columns at the left, we obtain the products shown in columns (6), (7), and (8). Summing the entries in these columns, we obtain the three values of d at the bottom of the table, -2.2, 3.2, and .9.

Summing the *squares* of the coefficients in columns (2), (3), and (4), of the table, we obtain the values of $\sum a_{.i}^2$ given at the bottom of the table. Thus, $\sum a_{.1}^2 = 2$, $\sum a_{.2}^2 = 2$, and $\sum a_{.3}^2 = 1$. From the analysis of variance

Table 10.5 Application of the Comparisons of Table 10.4 to the Means of the Analysis of Variance Given in Table 10.3

(1) Treatments	(2) Coefficients $a_{\cdot 1}$	(3) Coefficients $a_{\cdot 2}$	(4) Coefficients $a_{\cdot 3}$	(5) $\bar{X}_{k\cdot}$	(6) Products $a_{\cdot 1}\bar{X}_{k\cdot}$	(7) Products $a_{\cdot 2}\bar{X}_{k\cdot}$	(8) Products $a_{\cdot 3}\bar{X}_{k\cdot}$
1	1	0	½	17.2	17.2	0	8 6
2	−1	0	½	19.4	−19.4	0	9.7
3	0	−1	−½	15.8	0	−15.8	−7.9
4	0	1	−½	19.0	0	19.0	−9.5
$\sum a_{\cdot i}$	0	0	0	d	−2 2	3.2	.9
$\sum a_{\cdot i}^2$	2	2	1				

of Table 10.3, we have $s^2 = 3.06$. Then the standard errors, obtained by formula (10.2), for the three d values will be

$$s_{d_1} = \sqrt{\frac{3.06}{10}(2)} = .782$$

$$s_{d_2} = \sqrt{\frac{3.06}{10}(2)} = .782$$

$$s_{d_3} = \sqrt{\frac{3.06}{10}(1)} = .553$$

Dividing each d by its standard error, as in formula (10.3), we obtain three t's. Thus

$$t_1 = \frac{-2.2}{.782} = -2.813$$

$$t_2 = \frac{3.2}{.782} = 4.092$$

$$t_3 = \frac{.9}{.553} = 1.627$$

Each of these t's has 36 d.f., the number of degrees of freedom associated with s^2 of the analysis of variance. If $\alpha = .01$, and with a two-sided test, the first two comparisons, d_1 and d_2, are significant, whereas the third, d_3, is not.

Rules for Orthogonal Comparisons

Comparisons of the kind shown in Table 10.4 are called linear functions of the treatment means. Any linear function of the treatment means

$$d_i = a_{1i}\bar{X}_{1\cdot} + a_{2i}\bar{X}_{2\cdot} + \cdots + a_{ki}\bar{X}_{k\cdot} \qquad (10.4)$$

is called a comparison between the means, if the sum of the coefficients is equal to zero, that is, if $\sum a_{\cdot i} = 0$.

If a second comparison, d_j, is made, then d_i and d_j are said to be *orthogonal* or *independent*, if the sum of the products of the coefficients is equal to zero, that is, if

$$a_{1i}a_{1j} + a_{2i}a_{2j} + \cdots + a_{ki}a_{kj} = 0 \tag{10.5}$$

We note that the comparisons shown in Table 10.4 are *mutually orthogonal*, since the sum of the coefficients in each column is zero, and the sum of the products of the coefficients in each of the possible pairs of columns is also zero. For example, multiplying the coefficients in the first and second columns, we have $(1)(0) + (-1)(0) + (0)(-1) + (0)(1) = 0$. The sum of the products of the coefficients in the first and the third columns and the sum of the products of the coefficients in the second and third columns are also equal to zero.

ORTHOGONAL COMPARISONS OF TREATMENT SUMS

Instead of making our comparisons in terms of the treatment means, we may choose to use the treatment totals or sums. If n is the same for each group, and if d_1 is a comparison between the treatment means, then

$$d_1 = \frac{1}{n}(a_{11}\sum X_{1\cdot} + a_{21}\sum X_{2\cdot} + \cdots + a_{k1}\sum X_{k\cdot}) \tag{10.6}$$

Let the difference obtained by multiplying each of the treatment sums by the coefficients be D_1, that is let

$$D_1 = a_{11}\sum X_{1\cdot} + a_{21}\sum X_{2\cdot} + \cdots + a_{k1}\sum X_{k\cdot} \tag{10.7}$$

We shall refer to D_1 as a comparison between the *treatment sums*. Then it can also be shown that

$$A_1 = \frac{D_1^2}{n\sum a_{\cdot 1}^2} \tag{10.8}$$

will be a component of the treatment sum of squares or the sum of squares between groups with 1 d.f.

If a second comparison D_2 between the treatment sums is made, and if D_1 and D_2 are orthogonal, that is, if the sum of the products of the corresponding coefficients is zero, then

$$A_2 = \frac{D_2^2}{n\sum a_{\cdot 2}^2}$$

will be a component of the *residual treatment* sum of squares or a component of

$$\text{Residual} = \text{Treatments} - A_1$$

Similarly, after partitioning the treatment sum of squares into orthogonal components A_1 and A_2, we may then choose a comparison D_3 that is orthogonal to both D_1 and D_2. If all of the comparisons $D_1, D_2, \cdots, D_{k-1}$ are mutually orthogonal, that is, if every pair is orthogonal, then the sum of $A_1, A_2, \cdots, A_{k-1}$ will be equal to the treatment sum of squares. Thus

$$\text{Treatment sum of squares} = A_1 + A_2 + \cdots + A_{k-1}$$

where each of the A sums of squares has 1 d.f.

Test of Significance

Each of the A sums of squares may be tested for significance by finding

$$F = \frac{A}{s^2} \tag{10.9}$$

where s^2 is the error mean square of the analysis of variance. The F of formula (10.9) will have 1 d.f. for the numerator. The degrees of freedom for the denominator will be equal to those associated with s^2.

In Table 10.6 we give the treatment *sums* corresponding to the treatment means of Table 10.5. The coefficients given in Table 10.6 are the

Table 10.6 Application of the Comparisons of Table 10.4 to the Treatment Sums of the Analysis of Variance Given in Table 10.3

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatments	Coefficients			$\sum X_{k\cdot}$	Products		
	$a_{\cdot 1}$	$a_{\cdot 2}$	$a_{\cdot 3}$		$a_{\cdot 1}\sum X_{k\cdot}$	$a_{\cdot 2}\sum X_{k\cdot}$	$a_{\cdot 3}\sum X_{k\cdot}$
1	1	0	½	172	172	0	86
2	−1	0	½	194	−194	0	97
3	0	−1	−½	158	0	−158	−79
4	0	1	−½	190	0	190	−95
$\sum a_{\cdot i}$	0	0	0	D	−22	32	9
$\sum a_{\cdot i}^2$	2	2	1	D^2	484	1,024	81
$n\sum a_{\cdot i}^2$	20	20	10	A	24 20	51.20	8.10

same as those given in Table 10.4. The values of D are found by first multiplying the treatment sums by the coefficients given in the table to obtain the products in columns (6), (7), and (8). Summing the entries in these columns gives the values of D at the bottom of the table. Squaring the D values and dividing each by the corresponding value of $n\sum a_{\cdot i}^2$, we

obtain the A values or sums of squares. We note that $A_1 + A_2 + A_3 = 24.20 + 51.20 + 8.10 = 83.50$, the sum of squares between groups in the analysis of variance of Table 10.3.

Dividing each of the A sums of squares by $s^2 = 3.06$, we have, by formula (10.9),

$$F_1 = \frac{24.20}{3.06} = 7.91$$

$$F_2 = \frac{51.20}{3.06} = 16.73$$

$$F_3 = \frac{8.10}{3.06} = 2.65$$

With $\alpha = .01$, an F of 7.39 will be significant for 1 and 36 d.f. Thus, we would conclude that the comparisons D_1 and D_2 are significant, whereas D_3 is not.

The conclusions concerning significance reached by means of the F tests are exactly the same as those we arrived at by means of t tests for the same data. If we square the t's obtained previously for the same data, we may note that each t for a given comparison of the means, d_i, is equal, within rounding errors, to the corresponding value of F for the given comparison of the sums, D_i. For example, we have $t_1^2 = (-2\,813)^2 = 7.91$, $t_2^2 = (4.092)^2 = 16.74$, and $t_3^2 = (1.627)^2 = 2\,65$, and these values are, within rounding errors, equal to the corresponding F's.

That the F of formula (10.9) is exactly equal to the t^2 of formula (10.3) can easily be shown. Thus, for a given comparison d_i, we have

$$t^2 = \frac{d_i^2}{\frac{s^2}{n}\sum a_{\cdot i}^2} = \frac{\frac{1}{n^2}D_i^2}{\frac{s^2}{n}\sum a_{\cdot i}^2} = \frac{\left(\frac{n}{\sum a_{\cdot i}^2}\right)\left(\frac{1}{n^2}\right)D_i^2}{s^2} = \frac{A_i}{s^2} = F$$

Additional Points to Consider

Several points should be made with respect to orthogonal comparisons. In the first place, if we have a comparison D_i so that $\sum a_{\cdot i} = 0$, then we can multiply the coefficients for the comparison by any constant without changing the nature of the comparison. For example, in Table 10.6, if we multiplied each of the coefficients in column (4) by 2, then we would have 1, 1, -1, and -1. Multiplying each of the treatment sums by these new coefficients, we have

$$D_3 = (172 + 194) - (158 + 190) = 18$$

with $\sum a_{.3}^2 = (1)^2 + (1)^2 + (-1)^2 + (-1)^2 = 4$, and

$$A_3 = \frac{D_3^2}{n\sum a_{.3}^2} = \frac{(18)^2}{(10)(4)} = 8.1$$

as before. We would obtain exactly the same value for A_3, regardless of the value of the constant used in multiplying the coefficients in column (4). The value of D_3, of course, would be changed, but so would the value of $\sum a_{.3}^2$, with the result that we would obtain exactly the same value for A_3. The fact that we can multiply the coefficients for a given comparison by a constant, without changing the nature of the comparison or the value of A, is sometimes useful in simplifying the computations of the A sums of squares.

It is also possible to analyze the sum of squares between groups into more than one set of orthogonal comparisons. For example, with $k = 4$ means, the two *sets* of orthogonal comparisons shown below differ from each other and from the *set* of orthogonal comparisons of Table 10.5.

Treatments	Set 1			Set 2		
	$a_{.1}$	$a_{.2}$	$a_{.3}$	$a_{.1}$	$a_{.2}$	$a_{.3}$
1	−1	0	0	½	−½	−½
2	⅓	−1	0	½	½	½
3	⅓	½	−1	−½	−½	½
4	⅓	½	1	−½	½	−½

That each of the two sets of comparisons given above is orthogonal can easily be determined by showing that the sum of the coefficients in each column is zero and that the sum of the products of the pairs of coefficients in all possible pairs of columns in each set is also zero.

Since more than one set of orthogonal comparisons is possible for a given group of $k > 2$ means, which particular set of comparisons is to be made should be determined by experimental interests and planned at the same time the experiment is designed. The particular comparisons shown in Table 10.5, for example, might have been of experimental interest and planned in advance if the dependent variable was a measure of maze performance and if four groups of rats had been tested under the following conditions:

Group 1: a group tested after 12 hours of water deprivation
Group 2: a group tested after 24 hours of water deprivation
Group 3: a group tested after 12 hours of food deprivation
Group 4: a group tested after 24 hours of food deprivation

The first comparison of Table 10.5 would test for the difference between the 12 and 24 hour water deprived groups; the second comparison would test for the difference between the 12 and 24 hour food deprived groups; and the third comparison would test for the difference between the average performance of the water deprived and the food deprived groups.

Again, we emphasize that the methods of this and the previous section should be used only if the comparisons are orthogonal and if they have been planned in advance It may also be emphasized that we do not need to make all of the possible $k - 1$ orthogonal comparisons. In some cases the experimenter may only be interested in several of the possible comparisons. The methods of this section may be used for any orthogonal comparisons equal to or less than $k - 1$, where k is the number of treatment means.

The Case of Unequal *n*'s

If the number of observations is not the same for each of the several groups, then a linear function of the treatment *sums*

$$D_i = a_{1i}\sum X_1. + a_{2i}\sum X_2. + \cdots + a_{ki}\sum X_k.$$

is a comparison of the sums, if

$$a_{1i}n_1 + a_{2i}n_2 + \cdots + a_{ki}n_k = 0$$

and the divisor for ${D_i}^2$ will be

$$\sum n_i {a_i}^2 = n_1 {a_{1i}}^2 + n_2 {a_{2i}}^2 + \cdots + n_k {a_{ki}}^2$$

Then

$${A_i}^2 = \frac{{D_i}^2}{\sum n_i {a_i}^2}$$

will be a component of the sum of squares between groups with 1 d.f.

Two comparisons D_i and D_j are orthogonal if

$$a_{1i}a_{1j}n_1 + a_{2i}a_{2j}n_2 + \cdots + a_{ki}a_{kj}n_k = 0$$

If D_i and D_j are orthogonal, then

$$A_j = \frac{{D_j}^2}{\sum n_i {a_j}^2}$$

will also be a component of the treatment sum of squares with 1 d.f.

TREND ANALYSIS[4]

In some experiments, the treatments may consist of an ordered variable. For example, we might test different groups of rats after 0, 6, 12,

[4] Trend analysis is discussed in greater detail in Chapter 14.

18, ···, 36 hours of food or water deprivation. As another example, we might have different groups of subjects tested for retention after 1, 2, 3, 4, and 5 days. In other cases, the treatments may consist of increasing intensities of shock, of increasing amounts of reward, or of increasing amounts of practice. In still other experiments, the treatments may consist of increasing exposure times or of radar scopes varying in size.

If the treatments consist of an ordered variable and if we can assume that the differences between the treatments are uniform, that is, equal, then we may be interested in determining whether the treatment means (or sums) are functionally related to the different values of the treatment variable. We may, for example, be interested in finding out whether the treatment means are linearly related to the values of the treatment variable or whether they depart significantly from a linear relation. If the deviations

Table 10.7 Analysis of Variance for a Learning Experiment

Source of Variation	Sum of Squares	d f	Mean Square	*F*
Between trials	2,665.48	4	666.37	31.51
Within groups	951 73	45	21.15	
Total	3,617.23	49		

from linearity are significant, then we may wish to determine whether the trend of the means can be adequately described by a quadratic or second-degree equation.

Suppose, for example, that we have an experiment in which the dependent variable is a measure of learning and 5 groups of subjects have

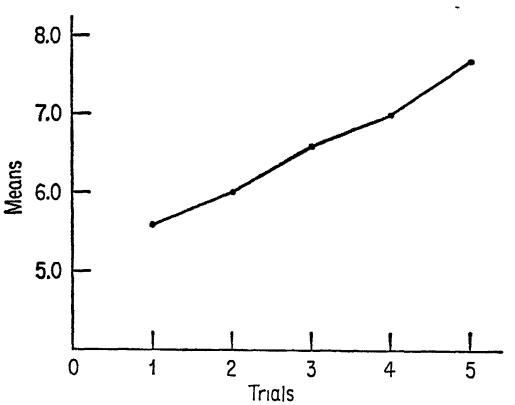


Figure 10.1 Mean learning scores plotted against trials. The sums for each trial are given in Table 10.8. The means are obtained by dividing each trial sum by $n = 10$.

been tested after 1, 2, 3, $\cdots$, 5 trials, respectively. With 10 subjects assigned to each group, we have the summary analysis of variance shown in Table 10.7. Table 10 8 gives the sum for each group and in Figure 10.1 the means, obtained by dividing the trial sums by $n = 10$, are plotted against the number of trials.

Orthogonal Polynomials

To test whether there is a significant linear trend, and also whether the means deviate significantly from linearity, we make use of a table of coefficients for orthogonal polynomials, Table XI in the Appendix.[5] This table gives the coefficients to be used in finding the linear and quadratic components of the treatment sum of squares. It may be observed that the coefficients in each row sum to zero and that, for any fixed value of k, the sum of the products of the coefficients for the linear and quadratic components is also zero. Thus, these coefficients meet the requirements for orthogonality discussed earlier.

For any given value of k, orthogonal coefficients corresponding to $k - 1$ components can be written. For example, if $k = 5$, the successive sets of coefficients would correspond to the linear, quadratic, cubic, and quartic components of the treatment sum of squares. Successive application of

Table 10.8 Orthogonal Coefficients for Linear ($a_{\cdot 1}$) and Quadratic ($a_{\cdot 2}$) Components for $k = 5$ Means

(1)	(2)	(3)	(4)	(5)	(6)
Trials	Coefficients		$\sum X_{k\cdot}$	Products	
	$a_{\cdot 1}$	$a_{\cdot 2}$		$a_{\cdot 1}\sum X_{k\cdot}$	$a_{\cdot 2}\sum X_{k\cdot}$
1	−2	2	564	−1,128	1,128
2	−1	−1	601	−601	−601
3	0	−2	663	0	−1,326
4	1	−1	703	703	−703
5	2	2	770	1,540	1,540
$\sum a_{\cdot i}^2$	10	14	D	514	38

these coefficients would enable one to determine how well the trend of the treatment means is represented by a polynomial of the first, second, third, and fourth degrees, respectively. Orthogonal coefficients for the higher degree polynomials can be found in Fisher and Yates (1948).

[5] The coefficients for orthogonal polynomials given in Table XI are for the case of equal intervals between the treatments. If the intervals are unequal, the coefficients given in Table XI should not be used. For procedures to be used with unequal intervals, see Grandage (1958) and Wishart and Metakides (1953).

Significance of a Linear Component

Using the *sums*, rather than the means, and multiplying these sums by the coefficients for the linear component, shown in column (2) and obtained from Table XI, we have

$$D_1 = (-2)(564) + (-1)(601) + (0)(663) + (1)(703) + (2)(770) = 514$$

Then $\sum a_{.1}^2 = 10$, and, since we have $n = 10$ observations for each sum, we find

$$A_1 = \frac{(514)^2}{(10)(10)} = 2{,}641.96$$

as the sum of squares for the linear component with 1 d.f. The sum of squares for deviations from linear regression will be equal to the sum of squares for trials minus the sum of squares for the linear component or 2,665.48 − 2,641.96 = 23.52, and this sum of squares will have $k - 2$ d.f.

Table 10.9 Test for Significance of Linear Regression and Deviations from Linear Regression for the Data of Table 10.8

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Linear regression	2,641.96	1	2,641.96	124.92
Deviations	23.52	3	7.84	
Within groups	951.73	45	21.15	
Total	3,617.21	49		

Table 10.9 summarizes the analysis. Testing the linear component for significance, we have, by formula (10.9),

$$F = \frac{2{,}641.96}{21.15} = 124.92$$

with 1 and 45 d.f. $F = 124.92$ is highly significant and we conclude that there is a significant linear trend in the trial means.

Significance of Deviations from Linearity

It is obvious that the mean square for deviations from the linear trend is not significant, since this mean square is 7.84 and the error mean square is 21.15. If the mean square for deviations from the linear trend is larger than the error mean square, we could test for its significance by finding[6]

$$F = \frac{\text{Mean square for deviations from linearity}}{\text{Error mean square}} \tag{10.10}$$

with $k - 2$ d.f. for the numerator.

[6] It is possible, but not in the present example, that the mean square for deviations from linearity is not significant, yet one of the mean squares for a higher-order polynomial is significant. Such cases would be rather unusual.

Significance of Curvature

If the F of formula (10.10) is significant, then we might also determine whether there is a significant curvature in the trend of the means by finding the quadratic component. Merely to illustrate the calculations, since we already know that the quadratic component cannot be significant, we multiply the trial sums by the coefficients for the quadratic component, as shown in column (3), to obtain

$$D_2 = (2)(564) + (-1)(601) + (-2)(663) + (-1)(703) + (2)(770) = 38$$

Then $\sum a_{.2}^2 = 14$ and

$$A_2 = \frac{(38)^2}{(10)(14)} = 10.31$$

with 1 d.f. If A_2 were larger than the error mean square, we would test it for significance by formula (10.9). If A_2 is significant this means that there is a significant curvature in the trend of the means.

DUNNETT'S TEST FOR COMPARISONS WITH A CONTROL

In some experiments the major objective is to compare each of a number of different treatments with a standard or control. Under these circumstances, the test procedure we shall use is one developed by Dunnett (1955). For example, in an experiment on the influence of incentives on learning, one group of subjects may be tested with a standard set of instructions and in the absence of any added incentives. This group may be designated the control group. The different treatments may then consist of varying kinds of incentives, introduced in an effort to improve performance over that of the control group. Our interest is in finding out which, if any, of the incentives result in performance significantly *better* than that of the control group.

Suppose, for example, that we have one control group and $k = 4$ treatment groups with $n = 8$ observations for each group.[7] We designate the mean of the control group by $\bar{X}_{0.}$ and the means of the treatment groups by $\bar{X}_{1.}, \bar{X}_{2.}, \cdots, \bar{X}_{k.}$. Each of the k treatment means is to be tested for significance by comparison with $\bar{X}_{0.}$. As we have stated above, our concern with the treatment means involves only the question of whether or not they are significantly *greater* than the control mean. We are not concerned with guarding against the alternative that a true treatment mean

[7] Dunnett (1955) indicates that the optimum allocation of subjects to the control and to each of the k treatment groups is approximately $n_0/n_1 = \sqrt{k}$, where n_0 is the number of observations for the control and n_1 the number for each of the k treatment groups. Thus, with $k = 4$ treatments we should have approximately twice as many subjects in the control group as in each of the treatment groups.

may be less than the true control mean. Table 10.10 gives the values of the dependent variable for the control group and $k = 4$ treatment groups.

Table 10.10 Values of a Dependent Variable for a Control Group and $k = 4$ Treatment Groups with $n = 8$ Observations for Each Group

Observations	Control	T_1	T_2	T_3	T_4
1	5	16	16	2	7
2	8	18	7	10	11
3	8	5	10	9	12
4	11	12	4	13	9
5	1	11	7	11	14
6	9	12	23	9	16
7	5	23	12	13	24
8	9	19	13	9	19
$\sum$	56	116	92	76	112
Means	7.0	14 5	11.5	9.5	14.0

Standard Error of a Comparison

Assuming that the variances for the five groups are all estimates of a common population variance, the estimate based upon the combined variances of the control and the k treatment groups will be the mean square within groups, with $(n - 1) + k(n - 1) = 35$ d.f. The mean square within groups, obtained in the usual way, is 24.17 for the observations in Table 10.10. Then the standard error of the difference between two means, as obtained from formula (7.16) will be

$$s_{\bar{x}_0 - \bar{x}_k} = \sqrt{\frac{(2)(24.17)}{8}} = 2.46$$

Table of Significant Values of t

Since we decided in advance that we were interested only in finding out whether the k treatment means exceed significantly the control mean, our tests of significance will be one-sided and we shall make k of them Table XIIa, in the Appendix, gives the values of t for a one-sided test with probability .95 of all k statements concerning the difference between a treatment mean and the control mean being correct. Table XIIc gives the corresponding values of t for the two-sided test, also with probability of .95 of all k statements concerning the differences being correct.[8] For the one-sided test, we enter Table XIIa with $k = 4$ and d.f. = 35. We have no entry for 35 d.f., but by interpolation between 30 and 40 we find $t = 2.24$.

[8] Tables XIIb and XIId give, respectively, the values of t for a one-sided and a two-sided test with probability .99 of all k statements concerning the difference between a treatment mean and the control mean being correct.

Tests of Significance

Instead of making successive t tests to determine whether the k differences between the treatment means and the control mean result in $t \geqq 2.24$, we solve for the magnitude of the difference itself that will be significant. In order for a difference to be declared significant, we must have

$$\frac{(\bar{X}_{k\cdot} - \bar{X}_{0\cdot}) - (m_k - m_0)}{2.46} \geqq 2.24$$

or, since the null hypothesis specifies that $m_k - m_0 = 0$,

$$\bar{X}_{k\cdot} - \bar{X}_{0\cdot} \geqq (2.46)(2.24)$$

or

$$\bar{X}_{k\cdot} - \bar{X}_{0\cdot} \geqq 5.51$$

Then any observed difference between a treatment mean and the control mean will be judged significantly greater than zero, if $\bar{X}_{k\cdot} - \bar{X}_{0\cdot} \geqq 5.51$. The observed differences are

$$\bar{X}_{1\cdot} - \bar{X}_{0\cdot} = 14.5 - 7.0 = 7.5 > 5.51$$

$$\bar{X}_{2\cdot} - \bar{X}_{0\cdot} = 11.5 - 7.0 = 4.5 < 5.51$$

$$\bar{X}_{3\cdot} - \bar{X}_{0\cdot} = 9.5 - 7.0 = 2.5 < 5.51$$

$$\bar{X}_{4\cdot} - \bar{X}_{0\cdot} = 14.0 - 7.0 = 7.0 > 5.51$$

and we conclude that the means for Treatments 1 and 4 are significantly greater than the mean of the control group and the remaining treatment means are not, with probability of .95 that these statements are all correct.

SCHEFFÉ'S TEST FOR MULTIPLE COMPARISONS

Scheffé (1953) has suggested a test that is appropriate for making *any* and *all* comparisons of interest between a set of k means, including those comparisons that may be suggested by the values of the means themselves. In other words, to use Scheffé's test *we do not need to plan the comparisons in advance.*[9] Table 10.11 shows various comparisons that might be made with respect to a set of $k = 4$ means, with n observations for each mean. Because of space limitations, we have entered the coefficients for the comparison in the rows of the table instead of the columns. Thus, for

[9] Scheffé's test, of course, can be used in testing planned orthogonal comparisons of the kind described earlier. His test will be more conservative than the procedures described for testing planned orthogonal comparisons; that is, larger differences will be required for significance. Scheffé suggests that with his test one might consider taking $\alpha = .10$ rather than $\alpha = .05$.

Table 10.11, *each row is a comparison between the treatment sums.* It is not necessary that all of the comparisons shown in the table be made, but we may make any that are of interest or all of them, using Scheffé's test, with probability equal to or greater than $1 - \alpha$ that all statements concerning significance are true. Thus, if $\alpha = .05$, the probability that all statements made will be correct will be $\geqq .95$.

Table 10.11 Possible Comparisons Between $k = 4$ Treatment Sums

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Comparison	$\sum X_{1.}$	$\sum X_{2.}$	$\sum X_{3.}$	$\sum X_{4.}$	$\sum a_{.i}^2$	D	D^2	A
	172	194	158	190				
1 vs. 2	1	−1	0	0	2	−22	484	24.20
1 vs. 3	1	0	−1	0	2	14	196	9.80
1 vs. 4	1	0	0	−1	2	−18	324	16.20
2 vs. 3	0	1	−1	0	2	36	1,296	64.80
2 vs. 4	0	1	0	−1	2	4	16	.80
3 vs 4	0	0	1	−1	2	−32	1,024	51.20
1 vs. 2 + 3	2	−1	−1	0	6	−8	64	1.07
1 vs. 2 + 4	2	−1	0	−1	6	−40	1,600	26.67
1 vs. 3 + 4	2	0	−1	−1	6	−4	16	.27
2 vs 1 + 3	−1	2	−1	0	6	58	3,364	56.07
2 vs. 1 + 4	−1	2	0	−1	6	26	676	11.27
2 vs 3 + 4	0	2	−1	−1	6	40	1,600	26.67
3 vs 1 + 2	−1	−1	2	0	6	−50	2,500	41.67
3 vs. 1 + 4	−1	0	2	−1	6	−46	2,116	35.27
3 vs. 2 + 4	0	−1	2	−1	6	−68	4,624	77.07
4 vs. 1 + 2	−1	−1	0	2	6	14	196	3.27
4 vs 1 + 3	−1	0	−1	2	6	50	2,500	41.67
4 vs 2 + 3	0	−1	−1	2	6	28	784	13.07
1 + 2 vs 3 + 4	1	1	−1	−1	4	18	324	8.10
1 + 3 vs 2 + 4	1	−1	1	−1	4	−54	2,916	72.90
1 + 4 vs. 2 + 3	1	−1	−1	1	4	10	100	2.50
1 vs. 2 + 3 + 4	3	−1	−1	−1	12	−26	676	5.63
2 vs. 1 + 3 + 4	−1	3	−1	−1	12	62	3,844	32.03
3 vs. 1 + 2 + 4	−1	−1	3	−1	12	−82	6,724	56.03
4 vs. 1 + 2 + 3	−1	−1	−1	3	12	46	2,116	17 63

Let any given comparison of the treatment sums be represented by D_i. The values of D, given in column (7) of Table 10.11, are obtained by multiplying each of the treatment sums, given at the top of the table, by the corresponding coefficients in each row. For a given D_i to be a comparison, we require only that $\sum a_{.i} = 0$ for the comparison, that is, for the

sum of the coefficients in the row to be zero It is not necessary for the various D comparisons to be mutually orthogonal.

Then, with each mean based upon the same number of observations, we have, by formula (10.8),

$$A_i = \frac{D_i^2}{n\sum a_{\cdot i}^2}$$

as the sum of squares for the comparison of interest. These sums of squares are given, for each row comparison, in column (9) of Table 10.11. Then, we may find

$$F = \frac{A_i}{s^2} \tag{10.11}$$

where s^2 is the error mean square of the analysis of variance.

Instead of evaluating the F of formula (10 11) in the usual way, by finding the tabled value for the 1 d f. corresponding to the numerator and the degrees of freedom associated with s^2 of the denominator, we compare it with the value of F'. We define F' as

$$F' = (k - 1)F \tag{10.12}$$

where F' is $k - 1$ times the tabled value of F for $k - 1$ and $k(n - 1)$ d.f. *F' is the standard in terms of which the F's of formula (10.11) are to be evaluated.*

From Table 10.3, we have $s^2 = 3.06$, with 36 d.f. We have $k = 4$ means and the tabled value of F for 3 and 36 d.f. is 2.86, with $\alpha = .05$. Then by formula (10.12) we have

$$F' = (4 - 1)(2.86) = 8.58$$

For any sum of squares, A_i, of Table 10.11, we know that $A_i/3.06$ must be equal to or greater than $F' = 8.58$ to be judged significant. Solving for the *smallest significant value of* A_i, we have

$$(F')(s^2) = (8.58)(3.06) = 26.25$$

and any A_i which equals or exceeds 26.25 in Table 10.11 will be judged significant.

Scheffé presents his test for comparisons between means rather than sums, that is, a comparison is given by d_i rather than by D_i. The standard error of d_i will be given by formula (10.1) or (10.2). Tests of significance can then be made by finding $t = d_i/s_{d_i}$, as given by formula (10.3). The t's thus obtained can be evaluated for significance by comparing them with $\sqrt{F'}$, that is, the square root of formula (10.12). Confidence limits for the d_i's can also be established in the manner described previously.

QUESTIONS AND PROBLEMS

In this chapter we shall give a limited number of problems. Additional problems, if they are desired, can be obtained by applying the methods of this chapter to the analysis of variance problems of other chapters and to the illustrative examples presented in the text.

1. We have a randomized groups design with $n = 6$ subjects assigned to each of 8 treatments. The error mean square of the analysis of variance is 53.02 with 40 d.f. The treatment means are given below:

A	*B*	*C*	*D*	*E*	*F*	*G*	*H*
19.7	36.7	50.6	51.4	55.1	61 3	65.3	72.0

Use Duncan's new multiple range test with $\alpha = .01$ to investigate the differences between the means.

2. Assume we have a control group and $k = 5$ treatment groups, with 10 observations for each group. The error mean square of the analysis of variance is 36.00. The means for the groups are given below:

Control	*A*	*B*	*C*	*D*	*E*
18.6	20.5	23.4	19.6	28.3	26.2

Use Dunnett's test to determine which of the treatment means is significantly greater than the mean of the control group.

3. For $k = 5$ treatment means, find a set of $k - 1$ orthogonal comparisons. Demonstrate that the comparisons are mutually orthogonal.

4. Describe an experiment involving 6 treatments in which a set of 5 planned and mutually orthogonal comparisons would be of experimental interest. Demonstrate that the comparisons are mutually orthogonal.

5. We have a randomized groups design in which the treatments consist of three equally spaced intervals of testing. One group is tested for retention of learned material after 12 hours, another after 24 hours, and the third after 36 hours. The means for the groups are 11.0, 9.0, and 5.0, respectively. We have $n = 10$ subjects in each group and $s^2 = 20.0$ with 27 d.f. (*a*) Use the analysis of variance to determine whether the means differ significantly. (*b*) Test the linear component of the trend of the means for significance.

6. Define, briefly, each of the following terms:

multiple comparisons
orthogonal comparisons
protection level
trend analysis

11

THE RANDOMIZED BLOCKS DESIGN

INTRODUCTION

In research in the behavioral sciences the experimental unit to which a treatment is applied is most often a subject—a person, a rat, a dog, a cat, a pigeon. It is well known, of course, that an unselected group of subjects will vary with respect to almost any variable that we might measure. Subjects differ in their reaction times, their ability to solve problems, to learn, to recall, to perceive, and so forth. In many experiments, the dependent variable may be one in which there are widespread individual differences. If, at the same time, the treatment effects are relatively slight, then an extremely large number of subjects may be required in each of the treatment groups in order to obtain a significant treatment mean square.

In this chapter we consider a design called a *randomized blocks* design. Under the circumstances described above, a randomized blocks design may be preferred to a randomized groups design. The randomized blocks design is based upon the principle of grouping experimental units into blocks. The blocks are formed with the hope that the units within each block will be more homogeneous in their response, in the absence of treatment effects, than units selected completely at random. By taking into account the differences existing between blocks in the analysis of variance, it is also anticipated that a smaller error mean square will be obtained, for the same number of observations, than if a randomized groups design had been used. This is the basis for preferring the randomized blocks design to the randomized groups design.

The term randomized blocks comes from agricultural experiments in which the experimental unit to which a treatment is applied is a plot of land. A block corresponds to a strip of land consisting of a number of adjacent or neighboring plots. It is often true that plots which are near one another in a field are more alike with respect to fertility and general soil condition than an equal number of randomly selected plots. The randomized blocks design, as used in agriculture, attempts to control for

some of the existing differences between randomly selected plots by grouping the plots into blocks.

In psychological research, the experimental unit corresponding to a plot is a subject. A group of subjects relatively homogeneous with respect to some variable corresponds to a block. In essence, each block of subjects is matched with respect to a given variable and for this reason the randomized blocks design in psychological research is also called a *matched groups* design. It is anticipated that each block of subjects will be relatively more homogeneous on the dependent variable in the absence of treatment effects than subjects selected completely at random.[1]

EXAMPLE OF A RANDOMIZED BLOCKS DESIGN

Suppose that the dependent variable in an experiment is the number of arithmetic problems correctly solved and that subjects are to be tested under $k = 5$ treatments Prior information available to the experimenter indicates that subjects vary considerably in the number of problems they can solve in a given period of time when tested under uniform conditions. The experimenter also has reason to believe that the treatment effects are apt to be slight. He therefore decides to use a randomized blocks design for his investigation.

Forming the Blocks

Let us assume that 25 subjects are to be used in the experiment with 5 subjects assigned to each of the 5 treatments. On the basis of an initial test, administered under uniform conditions, a score is obtained for each subject which represents the number of problems solved in a given period of time. These initial scores are used to arrange the subjects into blocks. In general, each block will consist of k subjects, where k is the number of treatments. With n such blocks available, we will have a total of kn observations. If the subjects are arranged in rank order of their scores on the initial test, then the first k subjects will make up the first block or group, the next k subjects the second block or group, and so on until n such blocks or groups have been formed. In the present example, we would have 5 blocks of 5 subjects each.

[1] In some experiments each subject is given every treatment, the order of the treatments being independently randomized for each subject. Each subject is thus regarded as a block in a randomized blocks design It is important to recognize that the analysis of variance for a randomized blocks design assumes that the treatment effects within a given block are independent of each other, that is, that there are no carry-over or residual effects from one treatment to another. If a single subject is administered a series of different treatments, it may be questionable as to whether we can assume that the treatment effects are, in fact, independent. This problem is discussed in greater detail in the chapter on Latin square designs.

Randomization

Within each block the subjects are assigned at random with one subject to each treatment. The randomization can be carried out by means of the sampling box described earlier. Since we have 5 subjects in each block, we put the disks with numbers 1, 2, 3, 4, and 5 in the box. We draw one disk and record the number. Without replacing the disk in the box, we shake the remaining disks, draw a second one, and record the number. We continue in this way until we have drawn 4 of the 5 disks, the number of the last disk being determined once we have selected 4 of the 5. This procedure will give a random permutation of the numbers from 1 to 5. We need one such random permutation for each block [2] Following the procedure described, the 5 random permutations shown in Table 11.1 were obtained. Now, if we have previously numbered the subjects from 1 to 5 in each of the 5 blocks, then Table 11.1 tells us which subject in each block is to be

Table 11.1 Random Permutations of the Subjects in Each Block in a Randomized Blocks Design with 5 Blocks and 5 Treatments

	Treatments				
	1	2	3	4	5
Block 1	4	3	5	1	2
Block 2	1	4	2	3	5
Block 3	2	1	5	3	4
Block 4	5	4	3	1	2
Block 5	2	1	4	5	3

assigned to which treatment. Thus, in Block 1, Subject 4 is to be assigned to Treatment 1. Let us assume that the measures obtained in the experiment are as given in Table 11.2.

Sums of Squares

The analysis of variance for the randomized blocks design begins with finding the total sum of squares. For the data of Table 11.2, we have

$$\text{Total} = (18)^2 + (17)^2 + \cdots + (16)^2 - \frac{(450)^2}{25} = 78.0$$

with $kn - 1$ d.f., where k is the number of treatments and n is the number of blocks.

[2] Similarly, we can use the table of random numbers to obtain random permutations. Suppose our point of entry in the table is block 02, row 02, and column 10. Reading down, the first random permutation we obtain is 5, 4, 3, 1 ,and 2. Continuing to read down the table we can obtain the other necessary random permutations.

Table 11.2 Observations in a Randomized Blocks Design with 5 Treatments and 5 Blocks

Block	Treatments					$\sum$	$\bar{X}_{.n}$	$\bar{X}_{.n} - \bar{X}_{..}$
	1	2	3	4	5			
1	18	20	20	21	21	100	20	2.0
2	17	19	19	20	20	95	19	1.0
3	16	17	18	19	20	90	18	.0
4	16	16	17	18	18	85	17	−1.0
5	16	16	15	17	16	80	16	−2.0
$\sum$	83	88	89	95	95	450		
$\bar{X}_{k}$	16.6	17.6	17 8	19 0	19.0	$\bar{X}_{..} = 18.0$		
$\bar{X}_{k.} - \bar{X}_{..}$	−1 4	−.4	− 2	1.0	1.0			

The sum of squares for treatments is found in the usual way. Thus,

$$\text{Treatments} = \frac{(83)^2}{5} + \frac{(88)^2}{5} + \cdots + \frac{(95)^2}{5} - \frac{(450)^2}{25} = 20.8$$

and the treatment sum of squares will have $k - 1$ d.f.

We now find the sum of squares for blocks and this will be given by

$$\text{Blocks} = \frac{(100)^2}{5} + \frac{(95)^2}{5} + \cdots + \frac{(80)^2}{5} - \frac{(450)^2}{25} = 50.0$$

The block sum of squares will have $n - 1$ d.f.

If we now subtract the treatment and block sums of squares from the total sum of squares we shall have a residual or remainder. Thus

$$\text{Residual} = \text{Total} - \text{treatments} - \text{blocks} \qquad (11.1)$$

which gives, for the present example,

$$\text{Residual} = 78.0 - 20.8 - 50.0 = 7.2$$

The degrees of freedom for the residual sum of squares may also be obtained by subtracting from the total degrees of freedom those for treatments and blocks. If we make this subtraction we have

$$(kn - 1) - (k - 1) - (n - 1) = kn - k - n + 1 = (n - 1)(k - 1)$$

as the degrees of freedom for the residual sum of squares.

Test of Significance

In Table 11.3 we show the sums of squares we have just calculated and the degrees of freedom associated with these sums of squares. Dividing

Table 11.3 Analysis of Variance of the Observations in Table 11.2

Source of Variation	Sum of Squares	d.f	Mean Square	F
Treatments	20.8	4	5.20	11.56
Blocks	50.0	4	12.50	
Residual	7 2	16	.45	
Total	78 0	24		

each sum of squares by its degrees of freedom, we obtain the mean squares given in the table. For the randomized blocks design, we have as a test of significance of the null hypothesis concerning the treatment means

$$F = \frac{\text{Treatment mean square}}{\text{Residual mean square}} \tag{11.2}$$

with $k - 1$ d.f. for the numerator and $(n - 1)(k - 1)$ d.f. for the denominator. In our example, we have $F = 11.56$ with 4 and 16 d.f. With $\alpha = .05$, $F = 11.56$ exceeds the tabled value of F and the null hypothesis would be rejected. We conclude that the treatment means do differ significantly.

Additional tests concerning the treatment means may be made in terms of procedures discussed previously under the heading multiple comparisons. For the randomized blocks design, s^2, the error mean square, is the residual mean square.

SUMS OF SQUARES IN THE RANDOMIZED BLOCKS DESIGN

In the randomized blocks design, the total sum of squares is analyzed into three component parts: the treatment sum of squares, the block sum of squares, and the residual sum of squares. The nature of the treatment sum of squares is already familiar. The block sum of squares is based upon the variation of the block means about the over-all mean. We have obtained the residual sum of squares by subtraction, but it can also be calculated directly.

For example, suppose we identify a given observation by X_{kn} where k represents a treatment and n a block.[3] Let k and n, when used as subscripts, be variables. Then, in the experiment described, k and n can take values from 1 to 5. Thus X_{32} would be the observation for Treatment 3 in Block 2. Let $X_{kn} - \bar{X}..$ represent a deviation from the over-all mean, $\bar{X}_{k.} - \bar{X}..$

[3] In a randomized groups design, if we should for one reason or another lose an observation for one of the treatments, the analysis of variance can still be used with unequal n's. With a randomized blocks design, however, the analysis of variance requires that we replace the missing value by an estimate. Methods for obtaining estimates of missing values can be found in Snedecor (1956), Federer (1955), Kempthorne (1952), and Cochran and Cox (1957).

the deviation of a treatment mean from the over-all mean, and $\bar{X}_{.n} - \bar{X}_{..}$ the deviation of a block mean from the over-all mean. Then, by subtraction, we have

$$(X_{kn} - \bar{X}_{..}) - (\bar{X}_{k.} - \bar{X}_{..}) - (\bar{X}_{.n} - \bar{X}_{..}) = X_{kn} - \bar{X}_{k.} - \bar{X}_{.n} + \bar{X}_{..}$$

and the right-hand side of the above expression represents a residual.

Table 11.4 shows the residuals for each observation in Table 11.2.[4] We note that the sum of the residuals in each column and in each row is zero. We have a total of kn residuals, but only $(n - 1)(k - 1)$ of them are

Table 11.4 Values of the Residuals $(X_{kn} - \bar{X}_{k.} - \bar{X}_{.n} + \bar{X}_{..})$ for the Observations in Table 11.2

Block	Treatments					Σ
	1	2	3	4	5	
1	−.6	.4	.2	.0	.0	.0
2	−.6	.4	.2	.0	.0	.0
3	− 6	−.6	.2	.0	1.0	.0
4	.4	−.6	.2	.0	.0	.0
5	1.4	.4	−.8	.0	−1 0	.0
Σ	0	0	.0	.0	.0	0

free to vary, since the sum of the residuals in each column and each row of the table is zero. If we square and sum the residuals in the table, we will obtain

$$(-.6)^2 + (-.6)^2 + \cdots + (-1.0)^2 = 7.2$$

or the residual sum of squares of the analysis of variance, as shown earlier in Table 11.3, with $(5 - 1)(5 - 1) = 16$ d.f.

Using the notation we have given, we can write the following general expression

$$X_{kn} - \bar{X}_{..} = (\bar{X}_{k.} - \bar{X}_{..}) + (\bar{X}_{.n} - \bar{X}_{..}) + (X_{kn} - \bar{X}_{k.} - \bar{X}_{.n} + \bar{X}_{..})$$

which states that the deviation of any given value from the over-all mean can be expressed as a sum of the three component parts on the right. If we square both sides of the expression and sum over all kn observations,

[4] In matrix notation it is customary to let the first subscript refer to the row variable and the second to the column variable. Thus in the manner in which we have presented the randomized blocks design in Table 11.2 the appropriate notation for a given observation would be X_{nk} rather than X_{kn}. The notation we have used would be in accord with general practice if we rearranged the data of Table 11.2 so that rows correspond to treatments and columns to blocks. This, however, is not convenient for data presentation. We have chosen, therefore, to maintain the form of presentation of Table 11.2 and to retain the notation introduced earlier.

we will find that all of the products between the terms on the right sum to zero.[5] Thus, we can write

$$\sum_{1}^{kn} (X_{kn} - \bar{X}..)^2 = n \sum_{1}^{k} (\bar{X}_{k\cdot} - \bar{X}..)^2 + k \sum_{1}^{n} (\bar{X}_{\cdot n} - \bar{X}.\,)^2 + \sum_{1}^{kn} (X_{kn} - \bar{X}_{k\cdot} - \bar{X}_{\cdot n} + \bar{X}..)^2$$

The term on the left will give the total sum of squares The first term on the right will give the treatment sum of squares and the second term the block sum of squares. The last term gives the residual sum of squares.

VARIABLES USED IN FORMING BLOCKS

In some cases it will be possible to obtain an initial measure, prior to the experiment proper, on the subjects to be used in the experiment with the same instrument which is to be used to measure the outcomes of the experiment. These initial measures may then be used to arrange the subjects into blocks. In other cases where it is not practical to obtain an initial measure on the same instrument, it may still be possible to group subjects into blocks on the basis of some other variable which we have reason to believe will tend to give us blocks that will be relatively homogeneous with respect to the dependent variable of interest. The variables used for forming blocks will depend, of course, upon the nature of the measurement made in the experiment. If we were studying the influence of various diets upon gain in weight, we might form blocks upon the basis of initial weights. In other experiments, the blocks may be formed on the basis of educational level, test performance, intelligence, age, and so forth. In rare cases, a block may consist of a pair of identical twins. The success of the randomized blocks design depends upon the degree to which subjects placed in the same block will, in fact, be relatively homogeneous in their performance on the dependent variable in the absence of treatment effects.[6] Consequently, we

[5] It may be helpful to write out the numerical values for a simple example. Take the following case:

Block 1	*A*	*B*	*C*	2	1	3
Block 2	*C*	*A*	*B*	6	5	1

Then, for the first observation in Block 1, we would have

$$-1.0 = (.5) + (-1.0) + (-.5)$$

If the 5 additional expressions are written, then it is easy to see that the products between the terms on the right sum to zero.

[6] If the distribution of measures used in forming the blocks is fairly normal, difficulties may be encountered in trying to form blocks for the two tails of the distribution, since the frequency of the extreme measures may not be sufficiently great to permit a

should not form blocks on the basis of variables which are irrelevant or unrelated to the measurements to be made in the experiment itself.

We have discussed the randomized blocks design as a means of reducing the error mean square of the analysis of variance, and we have considered a block as being formed on the basis of some prior information about the subjects to be used in the experiment. The randomized blocks design can also be used, however, to control for certain sources of variation that are not necessarily associated with individual differences between subjects. For example, suppose that an experiment involves 5 treatments and that it requires an experimental period of one hour for each treatment. Thus it may be possible to test only 5 subjects, one for each treatment, on a given day. Now, if there is substantial day-to-day variation, this source of variation could be controlled by regarding the 5 observations obtained in a single day as a block. In the analysis of variance, the block sum of squares would correspond to the day-to-day variation and would be eliminated from the error sum of squares.

NONADDITIVITY

A basic assumption of the randomized blocks design is the additivity of treatments over the range of blocks. We assume, for example, that the treatments will operate in the same way for each of the blocks. It can happen that this is not the case. Suppose, for example, that subjects are arranged into blocks on an initial measure of intelligence. The dependent variable is a measure of learning and several treatments are involved. Each of the blocks would correspond to a level of intellectual ability and we would have nonadditivity of treatments if one or more of the treatments operated differentially for different levels of intellectual ability.[7] It might

homogeneous grouping. On the other hand, if we restrict the blocks to those measures that are centrally located and for which the frequencies are the largest, then our treatments will be distributed over a less representative sampling of subjects. If we obtain the supplementary measures on many more subjects than we intend to actually use in the experiment, the problem of forming homogeneous and representative blocks will be considerably simplified.

[7] This condition is generally referred to as an *interaction* between blocks and treatments Previously, we pointed out that the experimenter may wish to be able to generalize about treatment effects over a representative sampling or range of the variable used in forming the blocks. Yet, if the blocks represent a wide range this may lead to an interaction between blocks and treatments. For this reason, Kempthorne (1955, pp. 964–965) has expressed the opinion that ". . . the requirement that the whole of the experimental material be as homogeneous as possible, which is the requirement in physical or chemical experiments, also holds for other experiments." The dilemma is that by restricting the range of differences between blocks, we may minimize the possibility of a block and treatment interaction, but at the same time we narrow the scope of any generalization about the significance of treatment effects. We may note

be true, for example, that one of the treatments tends to result in improved performance with subjects of high intellectual ability but lowered performance with subjects of low intellectual ability. Some other treatment may operate in the opposite manner. Perhaps still another treatment operates in such a way as to improve performance of those subjects with low intellectual ability but has no influence whatsoever on those subjects with high levels of intellectual ability. Any of these and various other conditions may result in nonadditivity.

Sum of Squares for Nonadditivity

A test for nonadditivity has been developed by Tukey (1949). We illustrate Tukey's test for the data of Table 11.2. We first find the deviations of each block mean from the over-all mean and each treatment mean from the over-all mean. These values are given in Table 11.2. We now multiply each observation in Block 1 of Table 11.2 by the corresponding value of $\bar{X}_{k\cdot} - \bar{X}_{\cdot\cdot}$ at the bottom of the table. For the first block, the sum of these products is

$$(18)(-1.4) + (20)(-.4) + (20)(-.2) + (21)(1.0) + (21)(1.0) = 4.8$$

and this sum of products is entered in column (1) of Table 11.5. The sum of products for each of the other blocks is obtained in the same way and entered in column (1) of Table 11.5. In column (2) of the table, we have

Table 11.5 Calculations for the Test for Nonadditivity for the Data of Table 11.2

Block	(1) $\sum_1^k (X_{kn})(\bar{X}_{k\cdot} - \bar{X}_{\cdot\cdot})$	(2) $(\bar{X}_{\cdot n} - \bar{X}_{\cdot\cdot})$	(3) $\sum_1^k (X_{kn})(\bar{X}_{k\cdot} - \bar{X}_{\cdot\cdot})(\bar{X}_{\cdot n} - \bar{X}_{\cdot\cdot})$
1	4.8	2.0	9.6
2	4.8	1.0	4.8
3	6.2	.0	.0
4	3.8	−1.0	−3.8
5	1.2	−2.0	−2 4
$\sum$	20.8	.0	8.2

entered the deviations of each block mean from the over-all mean. Column (3) gives the products of the entries in columns (1) and (2).

also that if all subjects are completely homogeneous, then nothing would be gained by using a randomized blocks design, since block differences on the dependent variable would be expected to be small and the error mean square of the randomized blocks design would be approximately that of a randomized groups design, but would have fewer degrees of freedom.

We now find $\sum_{1}^{k} (\bar{X}_{k\cdot} - \bar{X}_{\cdot\cdot})^2$ or the sum of squares of the deviations of the treatment means from the over-all mean Thus

$$\sum_{1}^{k} (\bar{X}_{k\cdot} - \bar{X}_{\cdot\cdot})^2 = (-1.4)^2 + (-.4)^2 + (-.2)^2 + (1.0)^2 + (1.0)^2 = 4.16$$

We also find $\sum_{1}^{n} (\bar{X}_{\cdot n} - \bar{X}_{\cdot\cdot})^2$ or the sum of squares of the deviations of the block means from the over-all mean. Thus

$$\sum_{1}^{n} (\bar{X}_{\cdot n} - \bar{X}_{\cdot\cdot})^2 = (2.0)^2 + (1.0)^2 + (.0)^2 + (-1.0)^2 + (-2.0)^2 = 10.0$$

The sum of squares for nonadditivity will be given by

$$\text{Nonadditivity} = \frac{\left[\sum_{1}^{kn} (X_{kn})(\bar{X}_{k\cdot} - \bar{X}_{\cdot\cdot})(\bar{X}_{\cdot n} - \bar{X}_{\cdot\cdot})\right]^2}{\sum_{1}^{k} (\bar{X}_{k\cdot} - \bar{X}_{\cdot\cdot})^2 \sum_{1}^{n} (\bar{X}_{\cdot n} - \bar{X}_{\cdot\cdot})^2} \qquad (11.3)$$

with 1 d.f. The value in the numerator of formula (11.3) is the sum of column (3) of Table 11.5 and is equal to 8.2. Substituting in the formula with the appropriate values, we have

$$\text{Nonadditivity} = \frac{(8.2)^2}{(4.16)(10)} = 1.62$$

Test of Significance

If we subtract the sum of squares for nonadditivity from the residual sum of squares, we obtain a *remainder*, as shown in Table 11.6. The mean

Table 11.6 Test for Nonadditivity for the Data of Table 11.2

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Nonadditivity	1.62	1	1.62	4 38
Remainder	5.58	15	.37	
Residual	7.20	16		

square for nonadditivity may be tested for significance by dividing it by the mean square for remainder. Thus

$$F = \frac{\text{Nonadditivity}}{\text{Remainder}} \qquad (11.4)$$

and the numerator will have 1 d.f. and the denominator $(n-1)(k-1)-1$ d.f. For the present example, we have $F = 1.62/.37 = 4.38$ with 1 and 15 d.f. With $\alpha = .05$, the tabled value of F is 4.54 and our obtained value is close to being significant.

Transformations and Discrepant Observations

If the F of formula (11.4) is significant, then we might examine the data to determine whether the nonadditivity is the result of our scale of measurement or whether it is the result of one or more unusually discrepant observations. In case no unusually discrepant observations are found, we might then consider possible transformations of the scale such that on the transformed scale we might have additivity. We may examine the two possibilities suggested in the following manner: first, we multiply the entries in each block by the corresponding deviations of the treatment means to obtain the products shown in column (1) of Table 11.5. We then plot the entries in column (1) against the corresponding *block means* as shown in Figure 11.1. Two s confidence limits may be established by finding

$$\text{Average of the sum of products} \pm 2\sqrt{\begin{pmatrix}\text{Sum of squares of}\\ \text{deviations of}\\ \text{treatment means}\end{pmatrix}\begin{pmatrix}\text{Mean square}\\ \text{for}\\ \text{remainder}\end{pmatrix}} \qquad (11.5)$$

In the example under consideration, the average of the sum of products of column (1) is $20.8/5 = 4.16$. Then since $\sum_{1}^{k} (\bar{X}_{k\cdot} - \bar{X}_{\cdot\cdot})^2 = 4.16$ and the mean square for the remainder is .37, we have

$$4.16 \pm 2\sqrt{(4.16)(.37)}$$

or

$$4.16 \pm 2.48$$

The limits, 1.68 and 6.64, are shown in Figure 11.1. An unusually discrepant observation will tend to be shown by a point in the figure that is high or low while the other points tend to fall along a horizontal line. Our obtained F for nonadditivity was of borderline significance and the fact that we have one point in the figure that is quite low indicates that we have an unusually discrepant observation.

A transformation of scale is indicated if the plotted points tend to follow a slanting regression line. In this case, a logarithmic or square root transformation may result in a scale on which the treatment effects are additive. The discussion by Tukey (1949) is of value in selecting an appropriate transformation.[8]

[8] See also the discussion by Moore and Tukey (1954).

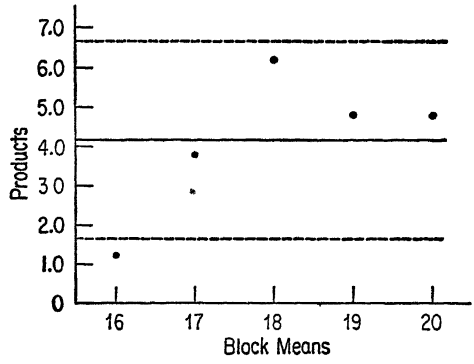


Figure 11.1 Plot of the products of column (1) of Table 11.5 against the block means given in Table 11.2. The solid line is the average of the sum of products and the broken lines correspond to 2s confidence limits.

RANDOMIZED BLOCKS WITH $k = 2$ TREATMENTS

Analysis of Variance

In Table 11.7 we give the results of a randomized blocks design in which we have $k = 2$ treatments. For the various sums of squares we have

$$\text{Total} = (14)^2 + (14)^2 + \cdots + (1)^2 - \frac{(340)^2}{40} = 424.0$$

$$\text{Treatments} = \frac{(180)^2}{20} + \frac{(160)^2}{20} - \frac{(340)^2}{40} = 10.0$$

$$\text{Blocks} = \frac{(28)^2}{2} + \frac{(26)^2}{2} + \cdots + \frac{(5)^2}{2} - \frac{(340)^2}{40} = 401.0$$

$$\text{Residual} = 424.0 - 10.0 - 401.0 = 13.0$$

The analysis of variance for the experiment is shown in Table 11.8. Testing the treatment mean square for significance, we have $F = 14.6$ with 1 and 19 d.f. With $\alpha = .05$, this is a highly significant value.

The t Test

In a randomized blocks design with $k = 2$ treatments, the test of significance of the treatment means can also be made in terms of the t test. Let $D = X_1 - X_2$ and these differences between the observations in each

Table 11.7 Observations in a Randomized Blocks Design with 2 Treatments and 20 Blocks

Block	Treatments		Σ	D
	1	2		
1	14	14	28	0
2	14	12	26	2
3	12	11	23	1
4	12	11	23	1
5	11	9	20	2
6	11	10	21	1
7	10	9	19	1
8	11	10	21	1
9	9	9	18	0
10	10	9	19	1
11	8	9	17	−1
12	10	9	19	1
13	8	8	16	0
14	7	8	15	−1
15	8	8	16	0
16	7	6	13	1
17	4	1	5	3
18	5	2	7	3
19	5	4	9	1
20	4	1	5	3
Σ	180	160	340	20

Table 11.8 Analysis of Variance of the Observations in Table 11.7

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Treatments	10.0	1	10.00	14.7
Blocks	401.0	19	21.11	
Residual	13.0	19	.68	
Total	424.0	39		

block are given in the last column of Table 11.7. Then the sum of squared deviations for the differences will be given by

$$\sum(D - \bar{D})^2 = \sum D^2 - \frac{(\sum D)^2}{n} \tag{11.6}$$

where n is the number of differences. For the differences in the table, we have

$$\sum(D - \bar{D})^2 = (0)^2 + (2)^2 + \cdots + (3)^2 - \frac{(20)^2}{20} = 26.0$$

The standard error of the difference between any two specified treatment means for the randomized blocks design will then be given by

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sum(D - \bar{D})^2}{n(n-1)}} \tag{11.7}$$

Substituting in formula (11.7), we obtain

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{26.0}{20(20-1)}} = .262$$

For $\bar{X}_1$. we have $180/20 = 9.0$ and for $\bar{X}_2$. we have $160/20 = 8.0$. Testing the null hypothesis $m_1 = m_2$, we have

$$t = \frac{9.0 - 8.0}{.262} = 3.82$$

with $n - 1 = 19$ d.f., where n is the number of differences or blocks. We note that $t^2 = (3.82)^2 = 14.6$ and t^2 is identical with the value of $F = 14.6$ for the randomized blocks design with $k = 2$ treatments in each block.[9]

Error Mean Squares in Randomized Blocks and Randomized Groups Designs

It can be shown that formula (11.7) is identical with

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2 - 2r_{12}s_{\bar{x}_1}s_{\bar{x}_2}} \tag{11.8}$$

[9] It should be clear that the t test can be used to test the difference between any two treatment means, even when we have $k > 2$ treatments in each block. If the assumptions of the analysis of variance for the randomized blocks design are satisfied, however, the analysis of variance is to be preferred, since our estimate of experimental error will be based upon a larger number of degrees of freedom. For example, if we have $k = 5$ treatments and $n = 10$ blocks, the estimate of experimental error based upon the analysis of variance will have 36 d.f. If we use the t test for the difference between a given pair of means, we shall be using an estimate of experimental error that is based upon 9 d.f.

where r is the correlation coefficient between the paired observations in the n blocks. With homogeneity of variance within treatments and with $n_1 = n_2 = n$, then formula (11.8) can be written

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{2s^2}{n}(1 - r_{12})} \tag{11.9}$$

It is obvious from formula (11.9) that the efficiency of the randomized blocks design, when $k = 2$, is dependent upon the degree of correlation present between the observations in the blocks. If the correlation is zero, then the standard error of formula (11.9) would be identical with that of a randomized groups design with two treatments.[10]

In general, for a randomized blocks design with any number of treatments, assuming homogeneity of variance within treatments, it can be shown that the residual mean square is related to the mean square within treatments in such a way that

$$\text{Residual} = s^2(1 - \bar{r}) \tag{11.10}$$

where s^2 is the mean square within treatments and $\bar{r}$ is the average intercorrelation of the $k(k - 1)/2$ possible values between the k treatment columns over the range of n blocks. If the average intercorrelation is positive, then the residual mean square will be less than the corresponding mean square within treatments for the *same* observations. An average intercorrelation of .50, for example, would result in a residual error mean square that is one-half the mean square within treatments. A negative $\bar{r}$ would, of course, give a residual mean square that is greater than the mean square within treatments for the *same* observations. It may be noted, however, that $\bar{r}$ can be -1.0 only if $k = 2$. The limiting negative value of $\bar{r}$ approaches zero as k increases. The limiting value is given by $-1/(k - 1)$. Thus with 5 treatments, the limiting negative value of $\bar{r}$ would be $-.25$.

QUESTIONS AND PROBLEMS

1. On the basis of an initial measure, subjects were formed into blocks of 3 subjects each. Treatments were assigned at random within each block. We have rearranged the observations within the blocks in accordance with the treatments.

[10] If the correlation is zero, we shall have identical standard errors but different degrees of freedom. For the randomized blocks design we would have only $n - 1$ d.f., whereas for the randomized groups design we would have $2(n - 1)$ d.f. for the test of significance.

Block	Treatment 1	Treatment 2	Treatment 3
1	21	20	22
2	20	19	21
3	20	19	22
4	18	19	20
5	18	17	18
6	18	18	19
7	18	16	19
8	16	15	18
9	16	13	15
10	15	14	16

Analyze the data using the analysis of variance. To investigate the difference between the various treatment means, it would be possible to use the procedures described previously under multiple comparisons.

2. Assume we have 5 treatments and a randomized blocks design with 6 blocks. Treatments were assigned at random within the blocks. We have rearranged the observations within the blocks in accordance with the treatments.

Block	Treatments					Sum
	A	*B*	*C*	*D*	*E*	
1	25	27	24	28	22	126
2	24	32	29	26	24	135
3	31	35	27	36	26	155
4	40	45	33	42	30	190
5	43	50	38	46	33	210
6	45	48	40	52	36	221
Sum	208	237	191	230	171	1,037

Analyze the data using the analysis of variance.

3. Assume that 20 subjects were pretested and then arranged into blocks with 2 subjects in each block. Treatments were assigned at random within blocks. We have rearranged the observations within the blocks in accordance with the treatments.

Block	Treatment 1	Treatment 2
1	2.5	3.6
2	4.6	5.7
3	9.3	8.9
4	4.5	6.7
5	1.5	1.9
6	6.4	7.8
7	4.7	4.6
8	5.6	5.9
9	7.3	6.9
10	6.6	7.0

(*a*) Analyze the data using the analysis of variance. (*b*) Analyze the same data using the t test of this chapter. You should find that $t^2 = F$.

4. We have a randomized blocks design with $k = 3$ treatments and $n = 30$ blocks. Treatments were assigned at random within each block. We have rearranged the observations within the blocks in accordance with the treatments.

Block	Treatments		
	A	*B*	*C*
1	18	19	15
2	15	14	13
3	16	18	17
4	18	14	15
5	19	13	13
6	20	16	11
7	19	17	15
8	17	16	14
9	18	17	13
10	20	16	14
11	16	16	19
12	11	15	18
13	11	19	19
14	16	17	21
15	15	15	18
16	14	16	21
17	13	16	20
18	16	15	17
19	14	14	18
20	14	17	19
21	17	17	15
22	13	15	19
23	16	18	18
24	12	15	17
25	13	15	17
26	15	18	18
27	13	16	16
28	13	16	18
29	14	16	15
30	14	14	17

Analyze the data using the analysis of variance.

5. Discuss the difference between a randomized groups design and a randomized blocks design.

12

THE 2 × 2 × 2 FACTORIAL EXPERIMENT

INTRODUCTION

Many experiments are concerned with the influence of two or more independent variables, usually called *factors*, on a dependent variable. The number of ways in which a factor is varied is called the number of *levels* of the factor. Thus, a factor which is varied in two ways would be said to have two levels and a factor which is varied in three ways would be said to have three levels. With two or more factors each with two or more levels, a treatment consists of a combination of one level for each factor. When the treatments consist of all possible different combinations of one level from each factor, and we have an equal number of observations for each treatment, the experiment is described as a *complete factorial experiment with equal replications.*[1]

A factorial experiment may be used with either of the two experimental designs we have discussed so far, the randomized groups design or the randomized blocks design, or with the experimental designs we shall discuss later. In this chapter we shall be concerned with the analysis of variance of a 2 × 2 × 2 factorial experiment. A 2 × 2 × 2 factorial experiment is one in which we have three factors with two levels for each factor. Although our discussion will be confined primarily to the 2 × 2 × 2 or 2^3 factorial, it can readily be generalized to any 2^n factorial experiment.

A 2 × 2 × 2 FACTORIAL EXPERIMENT

As an illustration, let us suppose that the dependent variable is a measure of the retention of verbal material. One factor of interest is the *number* of times the material is presented and this is varied in two ways by presenting the material once and by presenting the material twice. We shall designate this factor as A and the two levels as A_1, corresponding to

[1] In factorial experiments, using the methods of analysis described in this chapter, we should have equal n's for each treatment combination. For procedures dealing with unequal n's, see Snedecor (1956).

Table 12.1 Outcomes of a 2 × 2 × 2 Factorial Experiment with a Randomized Groups Design

	A_1				A_2			
	B_1		B_2		B_1		B_2	
	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2
	76	36	43	37	94	74	67	67
	66	45	75	22	85	74	64	60
	43	47	66	22	80	64	70	54
	62	23	46	25	81	86	65	51
	65	43	56	11	80	68	60	49
	43	43	62	27	80	72	55	38
	42	54	51	23	69	62	57	55
	60	45	63	24	80	64	66	56
	78	41	52	25	63	78	79	68
	66	40	50	31	58	61	80	58
$\sum$	601	417	564	247	770	703	663	556

TWO-PART ANALYSIS OF VARIANCE

Sums of Squares

We begin our analysis in a manner already familiar. We first find the total sum of squares, then the sum of squares between treatments, and the sum of squares within treatments. Thus

$$\text{Total} = (76)^2 + (66)^2 + \cdots + (58)^2 - \frac{(4{,}521)^2}{80} = 25{,}886.0$$

$$\text{Treatments} = \frac{(601)^2}{10} + \frac{(417)^2}{10} + \cdots + \frac{(556)^2}{10} - \frac{(4{,}521)^2}{80} = 19{,}507.9$$

$$\text{Within} = 25{,}886.0 - 19{,}507.9 = 6{,}378.1$$

As a check upon the arithmetic, we calculate the sum of squares within each of the 8 treatment groups. Then the sum of these sums of squares should be equal to the sum of squares within groups. For these 8 sums of squares we have,

$$\sum x_1^2 = (76)^2 + (66)^2 + \cdots + (66)^2 - \frac{(601)^2}{10} = 1{,}582.9$$

$$\sum x_2^2 = (36)^2 + (45)^2 + \cdots + (40)^2 - \frac{(417)^2}{10} = 590.1$$

$$\sum x_3^2 = (43)^2 + (75)^2 + \cdots + (50)^2 - \frac{(564)^2}{10} = 890.4$$

$$\sum x_4^2 = (37)^2 + (22)^2 + \cdots + (31)^2 - \frac{(247)^2}{10} = 402.1$$

$$\sum x_5^2 = (94)^2 + (85)^2 + \cdots + (58)^2 - \frac{(770)^2}{10} = 1{,}026.0$$

$$\sum x_6^2 = (74)^2 + (74)^2 + \cdots + (61)^2 - \frac{(703)^2}{10} = 576.1$$

$$\sum x_7^2 = (67)^2 + (64)^2 + \cdots + (80)^2 - \frac{(663)^2}{10} = 624.1$$

$$\sum x_8^2 = (67)^2 + (60)^2 + \cdots + (58)^2 - \frac{(556)^2}{10} = 686.4$$

Adding the above sums of squares, we have 1,582.9 + 590.1 + 890.4 + 402.1 + 1,026.0 + 576.1 + 624.1 + 686.4 = 6,378.1 for the sum of squares within groups. This is the same value we obtained by subtraction.

Homogeneity of Variance

Each of the sums of squares within each of the various treatment groups, when divided by the number of degrees of freedom, in this case 9, will provide a variance estimate s^2. Under the hypothesis that the population variance is the same for all treatment groups, the separate variances will all be estimates of the same parameter. Dividing each of the sums of squares by 9, we obtain as the separate estimates: 175.9, 65.6, 98.9, 44.7, 114.0, 64.0, 69.3, and 76 3. It may seem that these estimates vary quite a bit to be estimates of a common population variance and the experimenter may wish to test this null hypothesis before proceeding with the analysis of variance.

The manner of testing for homogeneity of variance has already been described and the test will not be repeated here with all of the calculations. It will suffice to say that the uncorrected value of the χ^2 for the present example is 5.91. Since we have 8 estimates, the number of degrees of freedom available for evaluating χ^2 will be equal to 7. From the table of χ^2 we find that $\chi^2 = 5.91$ with 7 d.f. has a probability of about .50 and there is no need to calculate the corrected value of χ^2. The data offer no significant evidence against the null hypothesis of random sampling from populations with the same variance.[3]

[3] As pointed out previously, care must be taken in interpreting a significant χ^2 in the test for homogeneity of variance, since the test is quite sensitive to nonnormality. It may be of some value, in case a significant χ^2 is obtained, to examine the frequency distribution of the residual deviations which, when squared, make up the error sum of squares.

Significance of the Treatment Mean Square

The analysis of variance up to this point has resulted in a partitioning of the total sum of squares and degrees of freedom into two parts. One part is associated with the differences between the 8 treatment groups and is based upon $k - 1 = 7$ d.f. The other part is associated with the variation within each of the treatment groups and has $k(n - 1) = 72$ d.f. This analysis is shown in Table 12.2. Testing the treatment mean square for

Table 12.2 Analysis of Variance Showing the Treatment Sum of Squares and the Sum of Squares Within Treatments for the Data of Table 12.1

Source of Variation	Sum of Squares	d.f	Mean Square	F
Treatments	19,507.9	7	2,786 8	31.5
Within treatments	6,378.1	72	88.6	
Total	25,886.0	79		

significance, we have $F = 2{,}786.8/88.6 = 31.5$ with 7 and 72 d.f. From the table of F we find that for 7 and 72 d.f., $F = 31.5$ is significant with probability less than .01. Thus we conclude that the treatment means differ significantly.[4]

PARTITIONING THE TREATMENT SUM OF SQUARES

Main Effects

In the experiment we have described, the treatment sum of squares has 7 d.f. We now consider a possible division of the treatment sum of squares into 7 component parts, each with 1 d.f. One of these components will be based upon a comparison of the sums for one and two presentations of the material and will be called the A sum of squares. Another will be based upon a comparison of the sums for the visual and auditory modes of presentation and will be called the B sum of squares. A third comparison will be based upon the sums for the immediate and delayed tests and will be called the C sum of squares. Each of these components will represent a comparison between the two levels of a given factor.

For the first comparison, we find the sum for $A_1 = 601 + 417 + 564 + 247 = 1{,}829$ and the sum for $A_2 = 770 + 703 + 663 + 556 =$

[4] Failure to obtain a significant difference between the treatment combinations is not necessarily a terminal test with a factorial design. Rather, the subsequent partitioning of the treatment sum of squares and tests of significance should be based upon the structure of the factorial design and the comparisons that have been planned. See the earlier discussion of orthogonal comparisons.

2,692. Each of these sums is based upon $(4)(10) = 40$ observations. Then the sum of squares for A will be given by

$$A = \frac{(1,829)^2}{40} + \frac{(2,692)^2}{40} - \frac{(4,521)^2}{80} = 9,309.6$$

For the second comparison, we have $B_1 = 601 + 417 + 770 + 703 = 2,491$ and $B_2 = 564 + 247 + 663 + 556 = 2,030$. Each of these sums is based upon $(4)(10) = 40$ observations. Then the sum of squares for B will be given by

$$B = \frac{(2,491)^2}{40} + \frac{(2,030)^2}{40} - \frac{(4,521)^2}{80} = 2,656.5$$

To find the sum of squares for C, we first find $C_1 = 601 + 564 + 770 + 663 = 2,598$ and $C_2 = 417 + 247 + 703 + 556 = 1,923$. Each of these sums is based upon $(4)(10) = 40$ observations. Then the sum of squares for C will be given by

$$C = \frac{(2,598)^2}{40} + \frac{(1,923)^2}{40} - \frac{(4,521)^2}{80} = 5,695.3$$

We have accounted for 3 of the 7 degrees of freedom associated with the treatment sum of squares. The A sum of squares corresponds to a comparison between A_1 and A_2 or between one and two presentations. The B sum of squares corresponds to a comparison between B_1 and B_2 or between the visual and auditory modes of presentation. The C sum of squares corresponds to a comparison between C_1 and C_2 or between the immediate and delayed tests. The comparisons between the levels of a factor are often called the *main effects* of the factors.

Interactions

In addition to the main effects of the factors, we have the possibility that there are *interactions* between the factors. We shall illustrate a method for calculating the interaction sum of squares between two factors when each factor has two levels, and postpone the discussion of the meaning of an interaction until later.

Table 12.3 Schematic Representation of the Two-Way Table for Computing an Interaction Sum of Squares with 1 d.f.

	B_1	B_2
A_1	a	b
A_2	c	d

The sum of squares for the interaction of A and B, designated $A \times B$, may be found by setting up a two-way table for the factors as shown in Table 12.3. Then the interaction sum of squares may be obtained by entering the sums corresponding to the cells of the table in the formula below.

$$\text{Interaction} = \frac{[(a + d) - (b + c)]^2}{(4)(n)} \tag{12.1}$$

where n is the number of observations contributing to each of the sums in the cells of the table. In the present problem each of the cell sums is based upon $(2)(10) = 20$ observations.

Table 12.4 gives the cell sums for the two-way tables for A and B, A and C, and B and C. For the two-way table for A and B, for example, the cell sums correspond to: a = the sum for $A_1B_1 = 601 + 417 = 1{,}018$; b = the sum for $A_1B_2 = 564 + 247 = 811$; c = the sum for $A_2B_1 = 770 + 703 = 1{,}473$; and d = the sum for $A_2B_2 = 663 + 556 = 1{,}219$. The cell sums for the other two-way tables have a similar interpretation.

Table 12.4 The Two-Way Tables for the $A \times B$, $A \times C$, and $B \times C$ Interactions

(a) Two-Way Table for A and B

	B_1	B_2	Σ
A_1	1,018	811	1,829
A_2	1,473	1,219	2,692
Σ	2,491	2,030	4,521

(b) Two-Way Table for A and C

	C_1	C_2	Σ
A_1	1,165	664	1,829
A_2	1,433	1,259	2,692
Σ	2,598	1,923	4,521

(c) Two-Way Table for B and C

	C_1	C_2	Σ
B_1	1,371	1,120	2,491
B_2	1,227	803	2,030
Σ	2,598	1,923	4,521

Substituting with the appropriate values from Table 12.4, we have as the $A \times B$ interaction sum of squares

$$A \times B = \frac{[(1{,}018 + 1{,}219) - (1{,}473 + 811)]^2}{(4)(20)} = \frac{(2{,}237 - 2{,}284)^2}{80} = 27.6$$

For the $A \times C$ interaction sum of squares we have

$$A \times C = \frac{[(1{,}165 + 1{,}259) - (664 + 1{,}433)]^2}{(4)(20)} = \frac{(2{,}424 - 2{,}097)^2}{80} = 1{,}336.6$$

and for the $B \times C$ interaction sum of squares we have

$$B \times C = \frac{[(1{,}371 + 803) - (1{,}120 + 1{,}227)]^2}{(4)(20)} = \frac{(2{,}174 - 2{,}347)^2}{80} = 374.1$$

Each of the interaction sums of squares we have just calculated will have 1 d.f. A general rule for determining the degrees of freedom associated with *any* interaction sum of squares is to multiply the degrees of freedom associated with the factors for which the interaction is being computed. Thus, in the present problem, we have 1 d.f. associated with A, 1 with B, and 1 with C, and the product of the degrees of freedom for A and B, for example, is 1 also.

The interaction sums of squares, $A \times B$, $A \times C$, and $B \times C$, each with 1 d.f., will account for 3 of the 4 degrees of freedom that were left after we found the sums of squares for the main effects of A, B, and C. The single remaining degree of freedom is associated with the sum of squares for the interaction of the three factors, designated by $A \times B \times C$. The degrees of freedom associated with the $A \times B \times C$ interaction sum of squares will be given by the products of the degrees of freedom associated with the factors involved in the interaction. Since we have but 1 d.f. for each of the factors, we will also have 1 d.f. for the $A \times B \times C$ interaction sum of squares.

The $A \times B \times C$ interaction sum of squares may be calculated directly and we will show methods for doing this later. In the present problem, it can be obtained most easily by subtraction. The sum of squares between treatments is equal to the sum of the sums of squares for A, B, C, $A \times B$, $A \times C$, $B \times C$, and $A \times B \times C$. Since we have already obtained all of these sums of squares except $A \times B \times C$, the latter can be obtained by subtracting the other 6 sums of squares from the treatment sum of squares. The sum of the 6 sums of squares calculated so far is equal to 19,399.7, and by subtraction from the treatment sum of squares, we obtain

$$19{,}507.9 - 19{,}399.7 = 108.2$$

as the sum of squares for the $A \times B \times C$ interaction.

Table 12.5 Complete Analysis of Variance for the Factorial Experiment of Table 12.1

Source of Variation		Sum of Squares	d.f.	Mean Square	F
A:	Number	9,309 6	1	9,309.6	105.07
B:	Mode	2,656.5	1	2,656.5	29.98
C:	Time	5,695.3	1	5,695.3	64.28
$A \times B$:	Number × Mode	27.6	1	27.6	
$A \times C$:	Number × Time	1,336.6	1	1,336.6	15.09
$B \times C$:	Mode × Time	374.1	1	374.1	4.22
$A \times B \times C$:	Number × Mode × Time	108.2	1	108.2	1.22
Error:	Within treatments	6,378.1	72	88.6	
	Total	25,886 0	79		

The summary of the complete analysis of variance is presented in Table 12.5, where we have divided the sums of squares by the number of degrees of freedom to obtain the mean squares. The values of F which have been entered in the table were obtained by dividing each of the mean squares which is to be tested for significance by the error mean square, that is, the mean square within treatments. Thus, each F in the table will be based upon 1 and 72 d.f. No value of F was calculated for the $A \times B$ interaction mean square since this mean square is obviously not significantly larger than the error mean square.

MEANING OF THE MAIN EFFECTS

From the table of F, we find that for 1 and 72 d.f., a value of F which is approximately equal to 4.0 will be significant at the 5 per cent level. For the main effects, A, B, and C, we have significant F's. The A mean square corresponds to a comparison between the means for one and two presentations *averaged over the two levels of B and the two levels of C*. The mean for one presentation or the first level of A can be obtained from Table 12.4 and is equal to 1,829/40 = 45.725. The mean for two presentations or the second level of A can also be obtained from Table 12.4 and is equal to 2,692/40 = 67.300. The fact that the A mean square is significant leads us to conclude that these two means differ significantly. Two presentations definitely result in a superior average retention compared with one presentation of the material.

Similarly, the main effect of B represents a comparison between the means for B_1, the visual mode, and B_2, the auditory mode, *averaged over the two levels of A and the two levels of C*. The mean for B_1 can be obtained from Table 12.4 and is equal to 2,491/40 = 62.275 and corresponds to the

mean for the visual mode of presentation. The mean for B_2 is equal to 2,030/40 = 50.750 and corresponds to the mean for the auditory mode of presentation. Since the mean square for B is significant in the analysis of variance, we conclude that the means for B_1 and B_2 differ significantly. The visual mode of presentation results in greater average retention than the auditory mode of presentation.

The main effect of C represents a comparison between the means for C_1, the immediate test, and C_2, the delayed test, *averaged over the two levels of A and the two levels of B*. These two means can be obtained from Table 12.4. The mean for C_1 is equal to 2,598/40 = 64.950 and corresponds to the mean for the immediate test The mean for C_2 is equal to 1,923/40 = 48.075 and corresponds to the mean for the delayed test. Since the C mean square of the analysis of variance is significant, we conclude that these two means differ significantly. We have greater average retention on the immediate test than on the delayed test.

THE INTERACTION EFFECTS

We come now to the interpretation of the interaction effects. Let us consider first the $A \times B$ interaction mean square which is *not* significant. The fact that this interaction mean square is not significant indicates that the difference between the means of A_1 and A_2 for the first level of B is not significantly different from the difference between the means of A_1 and A_2 for the second level of B. If the $A \times B$ interaction sum of squares were exactly zero, then the difference between the means of A_1 and A_2 for B_1 would be exactly equal to the difference between the means of A_1 and A_2 for B_2. With a nonsignificant $A \times B$ interaction, we can say that the A effect, the difference between A_1 and A_2, is *independent* of B, that is, we have approximately the same difference between A_1 and A_2, regardless of the levels of B.

We can see why, in order for the interaction sum of squares to be zero, what we have said above would have to be true. For example, setting the numerator of formula (12.1) equal to zero, we have

$$a - c = b - d$$

and this equality would have to hold if the interaction sum of squares is to be zero. The left-hand side of the above expression represents the difference between A_1 and A_2 for B_1, and the right-hand side the difference between A_1 and A_2 for B_2. If the $A \times B$ interaction mean square is significant, it means that the A effect is not the same for the different levels of B, that is, that $a - c$ and $b - d$ differ significantly.

Dividing each of the cell sums of Table 12.4(a) by 20, the number of observations contributing to the sums, we have as the mean difference between A_1 and A_2 for B_1

$$B_1: \qquad A_1 - A_2 = \frac{1{,}018}{20} - \frac{1{,}473}{20} = 50.90 - 73.65 = -22.75$$

and for the mean difference between A_1 and A_2 for B_2, we have

$$B_2: \qquad A_1 - A_2 = \frac{811}{20} - \frac{1{,}219}{20} = 40.55 - 60.95 = -20.40$$

and it is the fact that these two differences are much the same that results in a nonsignificant $A \times B$ interaction mean square.

Now, let us look at the $A \times C$ interaction mean square which is highly significant. Dividing each of the cell entries of Table 12.4(b) by 20, we have as the difference between the mean of A_1 and A_2 for C_1

$$C_1: \qquad A_1 - A_2 = \frac{1{,}165}{20} - \frac{1{,}433}{20} = 58.25 - 71.65 = -13.40$$

and as the difference between the means for A_1 and A_2 for C_2

$$C_2: \qquad A_1 - A_2 = \frac{664}{20} - \frac{1{,}259}{20} = 33.20 - 62.95 = -29.75$$

and we observe that these two differences are not at all comparable. Since the $A \times C$ mean square is significant, we know that the A effect is not independent of the C factor. In other words, the magnitude of the difference between A_1 and A_2 is not the same, within limits of random sampling, for C_1 and C_2. This is the meaning of the significant $A \times C$ interaction mean square.

Dividing each of the cell entries of Table 12.4(c) by 20, we have as the difference between the means of C_1 and C_2 for B_1

$$B_1: \qquad C_1 - C_2 = \frac{1{,}371}{20} - \frac{1{,}120}{20} = 68.55 - 56.00 = 12.55$$

and for the difference between the means of C_1 and C_2 for B_2

$$B_2: \qquad C_1 - C_2 = \frac{1{,}227}{20} - \frac{803}{20} = 61.35 - 40.15 = 21.20$$

and it is the failure of these two differences to be more alike that results in the significant $B \times C$ interaction.

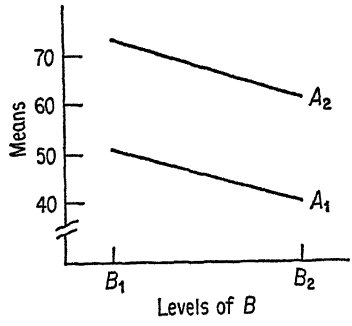


Figure 12.1 Means for levels of A at each level of B. A_1 and A_2 correspond to one and two presentations, respectively. B_1 and B_2 correspond to a visual and an auditory mode of presentation, respectively. Original data given in Table 12.4(a).

Another way of examining the nature of an interaction is to present it graphically. We take one of the factors, say B, for the X axis and graph the means for each level of A. For the example under discussion, the graphs for A_1 and A_2 are given in Figure 12.1. Each of the lines in the figure corresponds to a different level of A. If the lines for A_1 and A_2 were exactly parallel, then the $A \times B$ interaction would be zero. The fact that the lines are very nearly parallel, within the limits of random sampling, corresponds to the fact that the $A \times B$ interaction is not significant. Compare, however, the corresponding graphs for the $A \times C$ interaction, in Figure 12.2. Here we have taken C for the X axis and plotted the means for A_1 and A_2. Note that the lines for A_1 and A_2 are not parallel. The fact that the $A \times C$ interaction is significant is equivalent to stating that the lines A_1 and A_2 cannot be said to be parallel within the limits of random sampling.

Figure 12.3 gives the graph for the $B \times C$ interaction, where we have chosen B for the X axis. We have a significant $B \times C$ interaction mean square and this is shown graphically by the failure of the two lines, C_1 and C_2, in the figure to be parallel within the limits of random sampling.

The $A \times B \times C$ interaction mean square is not significant. But to examine the nature of the $A \times B \times C$ interaction, we consider the $A \times C$ interaction separately for each level of B, as shown in Table 12.6. The

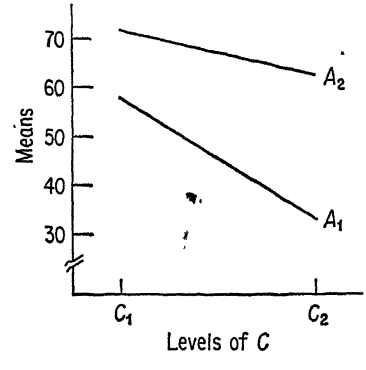


Figure 12.2 Means for levels of A at each level of C. A_1 and A_2 correspond to one and two presentations, respectively. C_1 and C_2 correspond to an immediate and a delayed test, respectively. Original data given in Table 12.4(b).

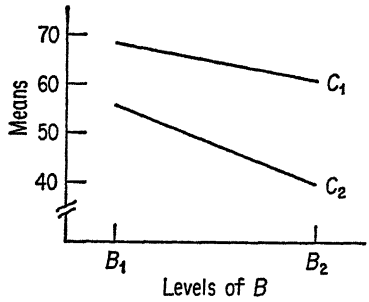


Figure 12.3 Means for levels of C at each level of B. C_1 and C_2 correspond to an immediate and a delayed test, respectively. B_1 and B_2 correspond to a visual and an auditory mode of presentation, respectively. Original data given in Table 12.4(c).

Table 12.6 Two-Way Table of Means for A and C for Each Level of B

	B_1		B_2		
	C_1	C_2		C_1	C_2
A_1	60.1	41.7	A_1	56.4	24.7
A_2	77.0	70.3	A_2	66 3	55.6

graphs for A_1 and A_2 against C for B_1 are shown in Figure 12.4(a) and the graphs for A_1 and A_2 against C for B_2 are shown in Figure 12.4(b).

Significance or lack of significance of a two factor interaction, $A \times C$ for example, tells us whether or not the A effect is the same for all levels

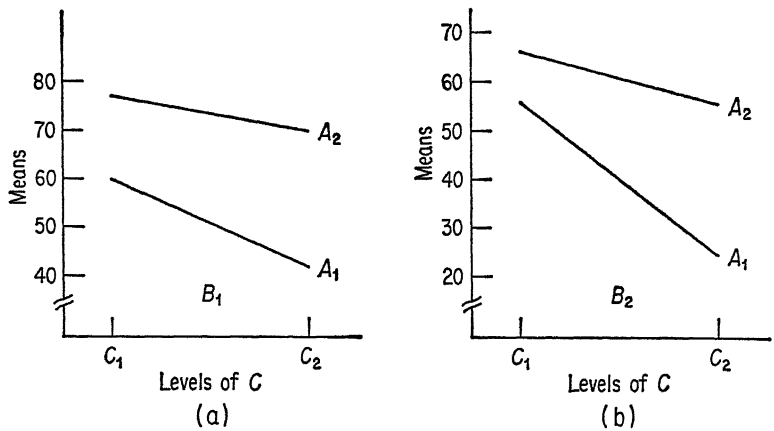


Figure 12.4 (*a*) Means for levels of A at each level of C for B_1. A_1 and A_2 correspond to one and two presentations, respectively. C_1 and C_2 correspond to an immediate and a delayed test, respectively. B_1 is a visual mode of presentation. Original data given in Table 12.6. (*b*) Means for levels of A at each level of C for B_2. A_1 and A_2 correspond to one and two presentations, respectively. C_1 and C_2 correspond to an immediate and a delayed test, respectively. B_2 is an auditory mode of presentation. Original data given in Table 12.6.

of C. Similarly, if we examine the $A \times C$ interaction separately for each level of B, and if these interactions are of the same form for each level of B, then the $A \times B \times C$ interaction will not be significant. A significant $A \times B \times C$ interaction, in other words, means that the $A \times C$ interaction is not the same for the different levels of B.

We note that the forms of the graphs in Figures 12.4a and 12.4b are fairly similar and this finding is consistent with the nonsignificance of the $A \times B \times C$ interaction mean square.

SUMMARY OF THE CONCLUSIONS

Let us summarize the conclusions based upon the analysis of variance for the experiment described. The significant A mean square tells us that the means for A_1 and A_2 averaged over the levels of B and C differ significantly. Examination of the means shows that two presentations are superior to one. The significant B mean square tells us that the means for B_1 and B_2 averaged over the levels of A and C differ significantly. Examination of these two means shows that the visual mode of presentation is superior to the auditory. The significant mean square for C tells us that the means for C_1 and C_2 averaged over the levels of A and B differ significantly. Examination of these two means shows that retention is greater on the immediate test than on the delayed test.

The $A \times B$ interaction is not significant. Therefore, the A effect, that is, the difference between A_1 and A_2, or between one and two presentations, is not dependent upon the particular mode of presentation employed. The $A \times B$ interaction is identical with the $B \times A$ interaction and our statement about the difference between A_1 and A_2 being independent of B is equivalent to stating also that the difference between B_1 and B_2 is independent of A.

We do have a significant $A \times C$ interaction. This tells us that the difference between A_1 and A_2 is not independent of the levels of C or, equivalently, that the difference between C_1 and C_2 is not independent of the levels of A. In other words, a statement about the A effect must be qualified by the particular level of C involved, or, equivalently, a statement about the C effect must be qualified by the particular level of A involved. The nature of the interaction can be shown by graphing the levels of C_1 and C_2 against those of A.

The interpretation of the significant $B \times C$ interaction with respect to the main effects of B and C is similar to that described above for the $A \times C$ interaction with respect to the main effects of A and C.

The $A \times B \times C$ interaction was discussed by considering the $A \times C$ interaction separately for each level of B. We could just as well have con-

sidered the $A \times B$ interaction separately for each level of C or the $B \times C$ interaction separately for each level of A. Just as the $A \times B$ interaction is a symmetrical property of A and B, so also the $A \times B \times C$ interaction is a symmetrical property for the three factors A, B, and C. The nonsignificance of the $A \times B \times C$ interaction, in other words, means that the $A \times B$ interactions for the separate levels of C are of the same form; that the $A \times C$ interactions for the separate levels of B are of the same form; and that the $B \times C$ interactions for the separate levels of A are of the same form.

If the $A \times B \times C$ interaction *is* significant, the nature of the interaction can be examined by the graphic methods presented earlier. Since the $A \times B \times C$ interaction is a symmetrical property of A, B, and C, we may consider graphing any one of the two factor interactions separately for the third factor. This is to say that we may examine the nature of the $A \times B \times C$ interaction by graphing the $A \times B$ interaction separately for the levels of C, or by graphing the $A \times C$ interaction separately for the levels of B, or by graphing the $B \times C$ interaction separately for the levels of A.

ORTHOGONAL COMPARISONS

The comparisons we have made with respect to the $2 \times 2 \times 2$ factorial experiment correspond to what we have described earlier as orthogonal comparisons, each with 1 d.f. That this is so can easily be seen from Table 12.7 where we show the nature of the comparisons in the manner described previously when we discussed orthogonal comparisons.[5] Because of space limitations, we have entered the coefficients in the rows of the table rather than the columns. We note, for example, that the sum of the coefficients in each row is zero and that the sum of products of the coefficient in each pair of rows is also zero. Multiplying the treatment sums by the corresponding coefficients, we obtain the comparisons shown in column D. Each treatment sum is based upon 10 observations and the sum of the squares of the coefficients in each row is 8. Then squaring D and dividing by $n\sum a_{.i}^2 = 80$, we have the sums of squares shown in column A. Each of these sums of squares has 1 d.f. and we note that they correspond exactly to the analysis of variance of Table 12.5.

The source of the coefficients for the main effects is apparent. The coefficients for a two-factor interaction are obtained by multiplying the corresponding coefficients for the factors involved in the interaction. For

[5] Although the numerators of the F ratios for orthogonal comparisons are independently distributed, the F ratios themselves are not. The reason for this is that for each ratio we have a common denominator or estimate of experimental error.

Table 12.7 Orthogonal Comparisons for the 2 × 2 × 2 Factorial

Comparison	Treatment Sums								$\sum a_i^2$	D	A
	$A_1B_1C_1$	$A_1B_1C_2$	$A_1B_2C_1$	$A_1B_2C_2$	$A_2B_1C_1$	$A_2B_1C_2$	$A_2B_2C_1$	$A_2B_2C_2$			
	601	417	564	247	770	703	663	556			
A	1	1	1	1	−1	−1	−1	−1	8	−863	9,309.6
B	1	1	−1	−1	1	1	−1	−1	8	461	2,656 5
C	1	−1	1	−1	1	−1	1	−1	8	675	5,695 3
$A \times B$	1	1	−1	−1	−1	−1	1	1	8	−47	27 6
$A \times C$	1	−1	1	−1	−1	1	−1	1	8	327	1,336 6
$B \times C$	1	−1	−1	1	1	−1	−1	1	8	−173	374.1
$A \times B \times C$	1	−1	−1	1	−1	1	1	−1	8	93	108.1

example, the coefficients for the $A \times B$ interaction are obtained by multiplying the coefficients in rows A and B and entering the product with the appropriate sign in row $A \times B$. To obtain the coefficients for the $A \times B \times C$ interaction, we multiply the corresponding coefficients in the rows for the A, B, and C comparisons. Thus the coefficients in row $A \times B \times C$ are obtained by multiplying the coefficients in rows A, B, and C.

In partitioning the treatment sum of squares into the 7 orthogonal comparisons, we assumed that the comparisons were the ones of experimental interest and that they were planned in advance. It would also be possible to examine specific comparisons between the treatment means or sums using Scheffé's test for multiple comparisons. In this way we could test certain comparisons that might be suggested by the data.

NOTATION AND SUMS OF SQUARES

We consider now a general notation for the factorial experiment. We let a = the number of levels of A, b = the number of levels of B, c = the number of levels of C, and n = the number of observations in each treatment group. The number of treatment groups will be $k = abc$ and the total number of observations will be kn. Then we let a general observation be X_{abcn} with the understanding that when a, b, c, and n are used as subscripts they represent variables. Thus with A at 2 levels, B at 2 levels, C at 2 levels, and with $n = 10$, a as a subscript can take values of 1 or 2, b can take values of 1 or 2, c can take values of 1 or 2, and n values of 1 to 10. Thus X_{1126} would correspond to the sixth observation of the first level of A, the first level of B, and the second level of C.

Then, using the dot notation, we can write

$$\begin{aligned} X_{abcn} - \bar{X}.... = {} & (X_{abcn} - \bar{X}_{abc\cdot}) \\ & + (\bar{X}_{a}... - \bar{X}....) \\ & + (\bar{X}_{\cdot b}.. - \bar{X}....) \\ & + (\bar{X}_{..c\cdot} - \bar{X}....) \\ & + (\bar{X}_{ab}.. - \bar{X}_{a}... - \bar{X}_{\cdot b}.. + \bar{X}....) \\ & + (\bar{X}_{a\cdot c\cdot} - \bar{X}_{a}... - \bar{X}_{..c\cdot} + \bar{X}....) \\ & + (\bar{X}_{\cdot bc\cdot} - \bar{X}_{\cdot b}.. - \bar{X}_{..c\cdot} + \bar{X}....) \\ & + (\bar{X}_{abc\cdot} + \bar{X}_{a}... + \bar{X}_{\cdot b}.. + \bar{X}_{..c\cdot} - \bar{X}_{ab}.. - \bar{X}_{a\cdot c\cdot} \\ & \quad - \bar{X}_{\cdot bc\cdot} - \bar{X}....) \end{aligned}$$

which states that the deviation of an observation from the over-all mean can be expressed as the sum of the eight terms on the right.

If we square both sides of the above expression and sum over all observations, we will find that the products of all terms on the right sum to zero. Thus, with $k = abc$, we can write

$$\begin{aligned}\sum_{1}^{kn}(X_{abcn} - \bar{X}....)^2 = \sum_{1}^{kn}(X_{abcn} - \bar{X}_{abc.})^2 \\ + bcn\sum_{1}^{a}(\bar{X}_{a...} - \bar{X}....)^2 \\ + acn\sum_{1}^{b}(\bar{X}_{.b..} - \bar{X}....)^2 \\ + abn\sum_{1}^{c}(\bar{X}_{..c.} - \bar{X}....)^2 \\ + cn\sum_{1}^{ab}(\bar{X}_{ab..} - \bar{X}_{a...} - \bar{X}_{.b..} + \bar{X}....)^2 \\ + bn\sum_{1}^{ac}(\bar{X}_{a.c.} - \bar{X}_{a...} - \bar{X}_{.c..} + \bar{X}....)^2 \\ + an\sum_{1}^{bc}(\bar{X}_{.bc.} - \bar{X}_{.b..} - \bar{X}_{..c.} + \bar{X}....)^2 \\ + n\sum_{1}^{k}(\bar{X}_{abc.} + \bar{X}_{a...} + \bar{X}_{.b..} + \bar{X}_{..c.} \\ - \bar{X}_{ab..} - \bar{X}_{a.c.} - \bar{X}_{.bc.} - \bar{X}....)^2\end{aligned}$$

The term on the left is the total sum of squares. The succeeding terms on the right give the sum of squares within treatments, the sum of squares for A, the sum of squares for B, the sum of squares for C, the $A \times B$ sum of squares, the $A \times C$ sum of squares, the $B \times C$ sum of squares, and the $A \times B \times C$ sum of squares.

FURTHER DISCUSSION OF INTERACTIONS

Consider the expression for any one of the two-factor interaction sums of squares. If we set up the two-way table involving these two factors, then the condition for a zero interaction is that each of the residuals of the table be equal to zero. That is, we will have a zero interaction for $A \times B$, for example, if

$$\bar{X}_{ab..} - \bar{X}_{a...} - \bar{X}_{.b..} + \bar{X}.... = 0$$

for all possible values.

This condition is met by the data of Table 12.8 where we show the table of means at the top and the corresponding residuals at the bottom of the table.

Table 12.8 Two-Way Table of Means for *A* and *B* with All Residuals Equal to Zero

	B_1	B_2	$\bar{X}$
A_1	5 0	10.0	7.5
A_2	15 0	20 0	17 5
$\bar{X}$	10 0	15.0	12.5

$$\bar{X}_{ab\cdot\cdot} - \bar{X}_{a\cdot\cdot} - \bar{X}_{\cdot b\cdot\cdot} + \bar{X}_{\cdot\cdot\cdot\cdot} = \text{Residual}$$

$$5\,0 - 7.5 - 10.0 + 12.5 = 0$$

$$10.0 - 7.5 - 15.0 + 12.5 = 0$$

$$15.0 - 17.5 - 10.0 + 12.5 = 0$$

$$20.0 - 17.5 - 15.0 + 12.5 = 0$$

A sufficient condition for a three-factor interaction sum of squares to be equal to zero, is that the tables of residuals for the two-factor interactions be exactly the same for each level of the third factor and hence the same as the table of residuals for the observations averaged over all levels of the third factor. This condition is met by the data of Table 12.9. There

Table 12.9 Two-Way Tables of Means for *A* and *B* for Each Level of *C* and Averaged over the Levels of *C* with Identical Residuals for Each Table

	C_1		$\bar{X}$
	B_1	B_2	
A_1	10.0	20.0	15.0
A_2	40 0	30.0	35.0
$\bar{X}$	25 0	25.0	25.0

Residuals		
	B_1	B_2
A_1	−5.0	5.0
A_2	5.0	−5.0

	C_2		$\bar{X}$
	B_1	B_2	
A_1	20 0	30.0	25.0
A_2	50.0	40.0	45.0
$\bar{X}$	35 0	35.0	35.0

Residuals		
	B_1	B_2
A_1	−5 0	5.0
A_2	5 0	−5.0

	$C_1 + C_2$		$\bar{X}$
	B_1	B_2	
A_1	15.0	25.0	20.0
A_2	45.0	35.0	40.0
$\bar{X}$	30.0	30.0	30.0

Residuals		
	B_1	B_2
A_1	−5 0	5.0
A_2	5.0	−5.0

we show the means for the $A \times B$ interactions separately for each level of C and also averaged over both levels of C. We note that the residuals for each of the three tables are identical. Thus, the $A \times B \times C$ interaction sum of squares will be equal to zero.

Table 12.10 gives the means for a $2 \times 2 \times 2$ factorial experiment where the $A \times B$ interaction *is not* zero, but where the $A \times B \times C$ interaction *is* equal to zero. Figure 12.5 shows the graphs of A_1 and A_2 against the levels of B separately for C_1 and C_2 and we observe that the forms of the two graphs are the same. The fact that the two graphs have the same form tells us that the $A \times B$ interaction is of the same form for C_1 and C_2. This, in turn, is equivalent to stating that the $A \times B \times C$ interaction is nonsignificant or, in the present case, zero Note also, in the lower figure, that when we average over C_1 and C_2, the lines of A_1 and A_2 have the same form as for each of the separate C levels.

Table 12.11 gives the means for a $2 \times 2 \times 2$ factorial experiment where the $A \times B$ interaction *is* zero, but where the $A \times B \times C$ interaction *is not* zero. Note that the forms of the graphs, as shown in Figure 12.6, for A_1 and A_2 against the levels of B separately for C_1 and C_2 are quite different. The fact that the two graphs differ in form tells us that the $A \times B$ interaction is not the same for C_1 as it is for C_2. Yet, when we average over C_1 and C_2, as in the lower figure, we see that the lines A_1 and A_2 are parallel indicating that the $A \times B$ interaction is zero. Thus, two-factor interactions, even when nonsignificant, must always be interpreted in accordance with whether or not the three-factor interaction is significant. If the three-factor interaction is significant, it means that the two-factor interactions are not the same for the different levels of the third factor. This can be true, as we have just seen, even when the two-factor interaction, averaged over the third factor, is zero or nonsignificant.

Table 12.10 Two-Way Tables of Means for A and B for Each Level of C and Averaged over the Levels of C with $A \times B \neq 0$ and $A \times B \times C = 0$

	C_1	
	B_1	B_2
A_1	10.0	20 0
A_2	40.0	30.0

	C_2	
	B_1	B_2
A_1	20.0	30.0
A_2	50 0	40.0

	$C_1 + C_2$	
	B_1	B_2
A_1	15 0	25.0
A_2	45.0	35.0

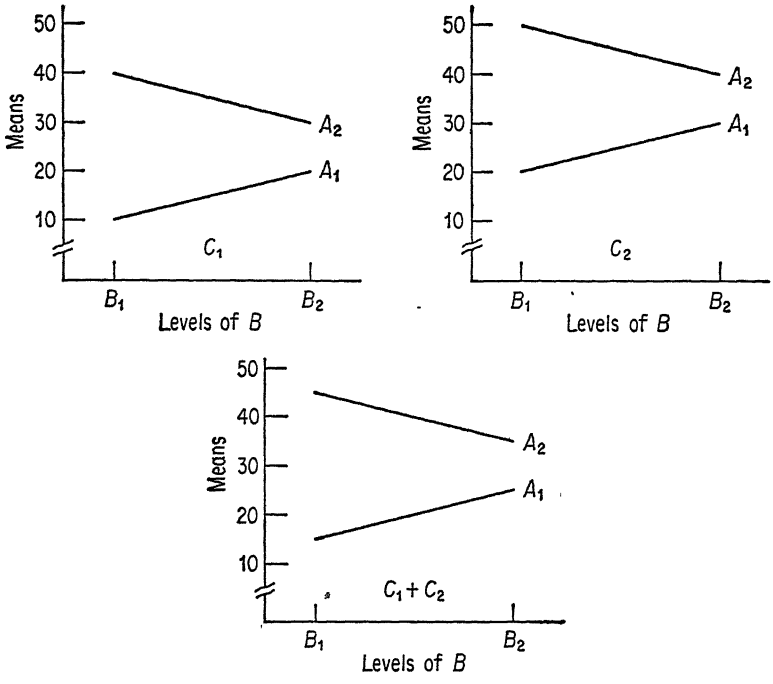


Figure 12.5 Means for levels of A at each level of B for C_1 and C_2 and averaged over the levels of C. The $A \times B \times C$ interaction is zero, but the $A \times B$ interaction is not zero. Original data given in Table 12.10.

Table 12.11 Two-Way Tables of Means for A and B for Each Level of C and Averaged over the Levels of C with $A \times B = 0$ and $A \times B \times C \neq 0$

	C_1	
	B_1	B_2
A_1	5.0	15.0
A_2	25.0	10.0

	C_2	
	B_1	B_2
A_1	5.0	5.0
A_2	5.0	30.0

	$C_1 + C_2$	
	B_1	B_2
A_1	5.0	10.0
A_2	15.0	20.0

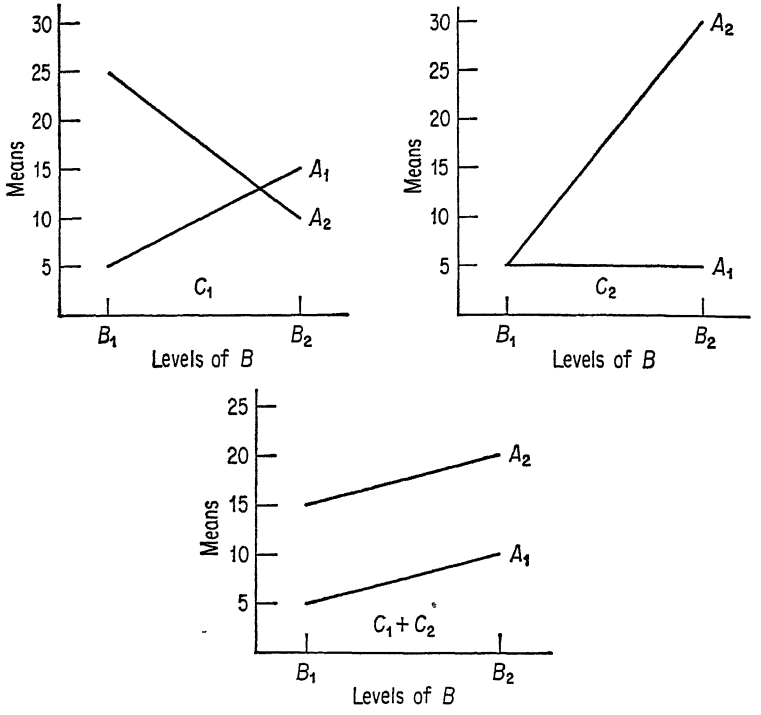


Figure 12.6 Means for levels of A at each level of B for C_1 and C_2 and averaged over the levels of C. The $A \times B \times C$ interaction is not zero, but the $A \times B$ interaction averaged over C_1 and C_2 is zero. Original data given in Table 12.11.

OTHER 2^n FACTORIAL EXPERIMENTS

We have considered only the $2 \times 2 \times 2$ factorial experiment. We could, of course, have a factorial experiment with only two factors or with more than three factors. With two factors, each varied in two ways, we would have 4 treatment groups, with 3 d.f. for the treatment sum of squares. The treatment sum of squares could then be analyzed into the sum of squares for A, the sum of squares for B, and the $A \times B$ interaction sum of squares, each with 1 d.f. The procedures for the analysis of variance of the 2×2 factorial experiment would be the same as those we have described for the $2 \times 2 \times 2$ factorial experiment.

If we have a factorial experiment with 4 factors, each varied in two ways, then we shall have 16 treatment groups with 15 degrees of freedom for the treatment sum of squares. To find the sums of squares for the main effects and interactions, we can make use of a table of orthogonal comparisons, as shown in Table 12.7 for the $2 \times 2 \times 2$ factorial. The coef-

ficients for the main effects are easy to enter in the table. Then the coefficients for the $A \times B$ interaction can be obtained by multiplying each of the coefficients in the row corresponding to the A effect by the corresponding coefficients in the row corresponding to the B effect. These products will give the coefficients for the $A \times B$ comparison. The coefficients for the $A \times C$, $A \times D$, $B \times C$, $B \times D$, and $C \times D$ interactions are obtained in the same manner. The coefficients for the $A \times B \times C$ interaction can be obtained by multiplying the corresponding coefficients of the A, B, and C comparisons. In the same manner, one can obtain the coefficients for any three-factor interaction and the coefficients for any four-factor interaction.

If we include more than four factors in an experiment, we shall be faced with the difficult problem of interpreting complex interactions of more than four factors, if they are found to be significant. As a general rule, it is suggested that we may be able to make better sense out of the results of our experiments if we do not attempt to include too many factors.

ADVANTAGES OF FACTORIAL EXPERIMENTS

The factorial experiment has a number of merits which may now be pointed out. It may be noted that, in the illustrative example, the full number of observations, that is, 80, entered into every comparison made, despite the fact that each treatment group contained but 10 observations. This is because, to take but one comparison—that between the number of presentations, the 80 observations could be split into a group of 40 which were alike in that they were based upon a single presentation of the material and another group of 40 which were alike in that they were based upon two presentations. The 40 observations in each of the two sets were not alike in other respects, but they were alike in sets of 10, differing only with respect to A_1 and A_2. For example, the treatment groups contributing to A_1 and A_2 are shown below:

A_1	A_2
$A_1B_1C_1$	$A_2B_1C_1$
$A_1B_1C_2$	$A_2B_1C_2$
$A_1B_2C_1$	$A_2B_2C_1$
$A_1B_2C_2$	$A_2B_2C_2$

Thus, we note that for each of these corresponding sets the experimental conditions are constant except for the levels of the one factor, A. In comparing A_1 and A_2 we are testing for the difference between these two levels averaged over levels of B and C that are the same for A_1 and A_2.

It should also be observed that the estimate of experimental error, the mean square within treatments, is based upon 72 d.f. If the experiment had been confined to a single factor with two levels and if an estimate of

experimental error is to be obtained with the same number of degrees of freedom, we would have to have 37 observations for each level of the factor or a total of 74 observations. And this experiment would provide information only about the single factor investigated. In the factorial experiment, on the other hand, with 80 observations, we not only have information about the main effects of three factors, but also about the interactions between these factors.

If the interactions involving a given factor are not significant, then we obviously have a broader basis for generalizing about the main effect of the factor, since it has been tested in conjunction with variation of other factors rather than holding the other factors constant at arbitrary levels. If, on the other hand, we have a significant two-factor interaction, examination of the interaction may provide us with additional insight as to how each factor operates.

Suppose, for example, we have a factorial experiment with two factors or drugs, A and B. Let A_1 and B_1 correspond to the *absence* of the drugs and A_2 and B_2 to the *presence* of the drugs in a standard dosage. Assume that each drug is supposedly a headache remedy and that the dependent variable is some measure of pain relief. If the $A \times B$ interaction is nonsignificant we have evidence that each drug operates independently. We may then make additional tests to determine whether there is a significant difference in the effectiveness of the two standard dosages and whether the administration of a combination of both drugs is superior to either drug alone. Methods for making these and various other tests have been described earlier under the heading of multiple comparisons.

If we have a significant $A \times B$ interaction, then we shall want to examine the nature of the interaction. We may find, for example, that each drug has a certain degree of effectiveness when present alone, but that the combination of the two drugs is no more effective than either drug alone. Or we may find that neither drug administered alone is effective, but that the combination of the two drugs is highly effective. It is also possible, of course, that the combination of both drugs may be less effective than either one administered alone.

The above discussion may serve to emphasize the theoretical and practical importance of examining interactions for whatever insight they may give us as to how a factor operates.[6]

[6] In a given experiment we may not be able to attach either theoretical or practical importance to the existing interactions. Under these circumstances we may be content with merely finding the rank order of the means for the treatment combinations and then, by means of Duncan's test, testing for differences within the set.

QUESTIONS AND PROBLEMS

1. The following data have been modified from an experiment by Glanville, Kreezer, and Dallenbach (1946). The problem was an investigation of the accuracy of apprehension of printed words under various experimental conditions. Three factors were selected for investigation: time of exposure, type size, and background. Let these factors be represented by A, B, and C, respectively, with each at two levels. We have A_1 corresponding to a 60-millisecond exposure and A_2 to a 120-millisecond exposure. We have B_1 corresponding to 6-point and B_2 to 12-point type. We also have C_1 corresponding to a blank and C_2 to a printed background.

Unfortunately, for our purposes, a list of only 100 words was used in the test conditions, and for the treatments with the longer exposure time the means were all close to the upper limit with the associated result of very small variances for these treatments. Not only is the variance heterogeneous, but, because the means for the treatments with the longer exposure time approach the upper limit, the distributions for these treatments were probably markedly skewed. A transformation of scale is probably in order. However, we shall do no violence to the conclusions arrived at by the experimenters if we assume slightly different experimental conditions than those actually used.

Let us assume that subjects were assigned at random to the 8 experimental conditions and that 50 subjects were used in each treatment group. Let us further assume that the lists contained more than 100 words and that the conditions of normality of distribution and homogeneity of variance are satisfied for all treatments. With these assumed conditions, we have the following sums (unchanged from the original experiment).

Exposure Time (in milliseconds)	Type (in points)	Background	Sum of Scores
60	6	Blank	1,319
120	6	Blank	4,592
60	6	Printed	1,196
120	6	Printed	4,365
60	12	Blank	3,682
120	12	Blank	4,939
60	12	Printed	3,357
120	12	Printed	4,885

The sum of squares within groups is given as equal to 84,397; the total sum of squares is given as equal to 405,084; and, by calculation, you will find the sum of squares between treatments is equal to 320,687. Complete the analysis of variance. If any interactions are significant, examine them by the graphic methods described in the chapter.

2. We have two factors, A and B, each at two levels. For each treatment combination, we have $n = 8$ subjects assigned at random. The data are as follows:

A_1		A_2	
B_1	B_2	B_1	B_2
8	5	10	5
6	8	9	7
9	10	4	3
9	7	8	5
8	10	8	3
7	7	4	5
6	8	3	5
3	5	6	8

Complete the analysis of variance for this 2×2 factorial.

3. In a $2 \times 2 \times 2$ factorial experiment we have the following measures:

A_1				A_2			
B_1		B_2		B_1		B_2	
C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2
8	5	10	5	7	6	5	2
6	8	9	7	10	8	7	7
9	10	4	3	6	7	4	5
9	7	8	5	7	6	7	7
8	10	8	3	5	8	6	5
7	7	4	5	7	9	8	9
6	8	3	5	6	8	10	6
3	5	6	8	10	9	6	6

Complete the analysis of variance.

4. Describe an experiment in which a significant two-factor interaction might be expected. Describe the nature of the interaction and state why it would be expected.

5. Consult a recent issue of the *Journal of Experimental Psychology* or some other journal which publishes the results of experiments. Find a study in which a significant two-factor interaction is reported. What is the nature of the interaction? Can you offer some explanation as to why it occurred?

6. Define, briefly, each of the following terms:

factor
levels of a factor
factorial experiment
main effect
interaction between two factors

13

FACTORIAL EXPERIMENTS: FURTHER CONSIDERATIONS

INTRODUCTION

A factorial experiment is not limited to the investigation of factors at only two levels, as in the examples cited in the last chapter. Factorial experiments may involve factors at several levels. If a factor has three or more levels, then the sum of squares for this factor will have more than 1 d.f. Then it also follows that the interactions of this factor with other factors will also have more than 1 d.f. The rule for determining the degrees of freedom for an interaction sum of squares, as stated previously, is to find the product of the degrees of freedom associated with the factors involved in the interaction.

Since the method of calculating the sum of squares for a two-factor interaction, when the interaction has more than 1 d.f., differs somewhat from the methods previously described, we shall examine a factorial experiment which involves interactions with more than 1 d.f. We shall also show methods for the direct calculation of any interaction, regardless of the number of factors or the number of levels of the factors involved in the interaction. It should be possible, then, for the reader to generalize from the examples described to any particular factorial experiment in which he is interested.

A $4 \times 3 \times 2$ FACTORIAL EXPERIMENT

Let us take as an example an experiment in which three factors are involved, namely, A, B, and C. Suppose that A has 4 levels, B has 3 levels, and C has 2 levels. Then we shall have $(4)(3)(2) = 24$ different treatments. We shall assume that a randomized groups design is used, with $n = 5$ observations for each treatment, so that we have a total of 120 obser-

vations.[1] Then the total sum of squares with 119 d.f. can be partitioned into the following components:

Main effects:	A	3
	B	2
	C	1
Two-factor interactions:	$A \times B$	6
	$A \times C$	3
	$B \times C$	2
Three-factor interaction:	$A \times B \times C$	6
Error:	Within treatments	96

Calculation of the Sums of Squares

The method of calculating the within-treatments sum of squares would be exactly the same as in the examples previously described. We could calculate the total sum of squares and the sum of squares between treatments and obtain the sum of squares within treatments by subtraction. Or we could calculate the sum of squares within each treatment group separately and the sum of these sums of squares would be equal to the sum of squares within treatments. Let us suppose that the within-treatments sum of squares has already been calculated and has been found to be equal to 1,198.00, and that we have added together the values of the observations for each treatment to obtain the sums entered in Table 13.1. Each sum in

Table 13.1 Outcomes of a 4 × 3 × 2 Factorial Experiment—Each Cell Entry Is the Sum of $n = 5$ Observations

		A_1	A_2	A_3	A_4	$\sum$
C_1	B_1	60	90	94	86	330
	B_2	54	92	98	96	340
	B_3	70	76	80	60	286
C_2	B_1	58	72	78	84	292
	B_2	76	82	74	64	296
	B_3	66	56	72	78	272
	$\sum$	384	468	496	468	1,816

the table is based upon $n = 5$ observations. Then the sum of squares between treatments, for which we have the treatment sums in the table, will be given by

$$\text{Treatments} = \frac{(60)^2}{5} + \frac{(54)^2}{5} + \cdots + \frac{(78)^2}{5} - \frac{(1{,}816)^2}{120} = 783.47$$

[1] Each treatment combination should have the same number of observations, if the methods of analysis described in this chapter are to be used.

It is this sum of squares, 783.47, based upon 23 d.f., that is to be analyzed into the sums of squares for the main effects and interactions.

From the data of Table 13.1 we may set up a table for factors A and C summed over the levels of B. Thus we obtain Table 13.2. Since B has 3

Table 13.2 The $A \times C$ Table with the Cell Entries Summed over the Levels of B—Each Cell Entry Is the Sum of 15 Observation

	A_1	A_2	A_3	A_4	Σ
C_1	184	258	272	242	956
C_2	200	210	224	226	860
Σ	384	468	496	468	1,816

levels, each sum in the table will be based upon $(3)(5) = 15$ observations. The two sums, 956 and 860, are the sums for C_1 and C_2, respectively. Each of these two sums is based upon $(4)(15) = 60$ observations. Then, as the sum of squares for C, we have

$$C = \frac{(956)^2}{60} + \frac{(860)^2}{60} - \frac{(1{,}816)^2}{120} = 76.80$$

The sum of squares for A will be based upon the sums for A_1, A_2, A_3, and A_4, given at the bottom of Table 13.2. Each of these sums is based upon $(2)(15) = 30$ observations, and the sum of squares for A will be

$$A = \frac{(384)^2}{30} + \frac{(468)^2}{30} + \frac{(496)^2}{30} + \frac{(468)^2}{30} - \frac{(1{,}816)^2}{120} = 235.20$$

The $A \times C$ interaction sum of squares may also be obtained from Table 13.2. We calculate first the sum of squares between the eight sums entered in the cells of the table, keeping in mind that each of these sums is based upon 15 observations. Thus,

$$\text{Between cells} = \frac{(184)^2}{15} + \frac{(200)^2}{15} + \cdots + \frac{(226)^2}{15} - \frac{(1{,}816)^2}{120} = 405.87$$

Then the $A \times C$ interaction sum of squares may be obtained by subtracting the sum of squares for A and the sum of squares for C, which we have already calculated, from the sum of squares between cells. As a general formula, if we have a two-way table with rows corresponding to the levels of one factor and columns corresponding to the levels of a second factor, the interaction sum of squares between the row and column factors will be given by

$$\text{Between cells} - \text{rows} - \text{columns} \qquad (13.1)$$

Then, remembering that the $R \times C$ interaction is the same as the $C \times R$ interaction, we have as the $A \times C$ interaction sum of squares

$$A \times C = 405.87 - 76.80 - 235.20 = 93.87$$

We now go back to the data of Table 13.1 and set up another two-way table for factors A and B summed over the levels of C. In this way we obtain

Table 13.3 The $A \times B$ Table with the Cell Entries Summed over the Levels of C—Each Cell Entry Is the Sum of 10 Observations

	A_1	A_2	A_3	A_4	$\sum$
B_1	118	162	172	170	622
B_2	130	174	172	160	636
B_3	138	132	152	138	558
$\sum$	384	468	496	468	1,816

the entries in Table 13.3. Since C has 2 levels, each sum in the table will be based upon $(2)(5) = 10$ observations. The sums for B_1, B_2, and B_3 are the row sums, 622, 636, and 558, respectively. Each of these sums is based upon $(4)(10) = 40$ observations. Then for the B sum of squares we have

$$B = \frac{(622)^2}{40} + \frac{(636)^2}{40} + \frac{(558)^2}{40} - \frac{(1,816)^2}{120} = 86.47$$

To obtain the $A \times B$ interaction sum of squares, we first calculate the sum of squares between the cells of Table 13.3. Thus

$$\text{Between cells} = \frac{(118)^2}{10} + \frac{(130)^2}{10} + \cdots + \frac{(138)^2}{10} - \frac{(1,816)^2}{120} = 425.87$$

We have already calculated the A (column) sum of squares and the B (row) sum of squares so that, by substitution in formula (13.1), we have

$$A \times B = 425.87 - 86.47 - 235.20 = 104.20$$

To find the $B \times C$ interaction sum of squares, we set up still another two-way table for factors B and C summed over the levels of A. Thus, from Table 13.1, we obtain Table 13.4. Since A has 4 levels, each sum in the table will be based upon $(4)(5) = 20$ observations. For this table we have already calculated the C (row) sum of squares and the B (column)

Table 13.4 The $B \times C$ Table with the Cell Entries Summed over the Levels of A—Each Cell Entry Is the Sum of 20 Observations

	B_1	B_2	B_3	$\sum$
C_1	330	340	286	956
C_2	292	296	272	860
$\sum$	622	636	558	1,816

sum of squares. Thus, all that we need to find is the sum of squares between cells and we can then obtain the $B \times C$ interaction sum of squares by formula (13.1). For the sum of squares between cells we have

$$\text{Between cells} = \frac{(330)^2}{20} + \frac{(292)^2}{20} + \cdots + \frac{(272)^2}{20} - \frac{(1{,}816)^2}{120} = 175.87$$

Then, by subtraction, we have

$$B \times C = 175.87 - 76\,80 - 86.47 = 12.60$$

We shall show how to calculate the $A \times B \times C$ interaction sum of squares later. For the moment, we shall obtain it by subtraction. We know that the sum of squares for A, B, C, $A \times B$, $A \times C$, $B \times C$, and $A \times B \times C$ must equal the treatment sum of squares. Since we have calculated the first 6 of these 7 sums of squares, we can obtain the last one by subtraction. The sum of the sums of squares we have calculated is equal to $235.20 + 86.47 + 76\,80 + 104.20 + 93.87 + 12.60 = 609.14$. Subtracting this sum from the treatment sum of squares, we have

$$A \times B \times C = 783.47 - 609.14 = 174.33$$

Summary of the Analysis

We have assumed that the within-treatments sum of squares has already been calculated and found to be equal to 1,198.00 The sums of

Table 13.5 Analysis of Variance of the 4 × 3 × 2 Factorial Experiment of Table 13.1

Source of Variation	Sum of Squares	d f.	Mean Square	F
A	235.20	3	78.40	6.28*
B	86.47	2	43.24	3.46*
C	76.80	1	76.80	6.15*
$A \times B$	104.20	6	17.37	1.39
$A \times C$	93.87	3	31.29	2.51
$B \times C$	12.60	2	6.30	
$A \times B \times C$	174.33	6	29.06	2.33*
Within treatments	1,198.00	96	12.48	
Total	1,981.47	119		

squares we have just calculated and the within-treatments sum of squares are shown in Table 13.5, which summarizes the analysis.[2]

[2] If we have factors at more than two levels, it is possible to obtain a set of orthogonal comparisons separately between the levels of each factor, such that each comparison has 1 d.f. Thus, for example, if we have three levels for A and three for B, we can obtain two orthogonal comparisons between the levels of each factor. If we then multiply the coefficients for a given orthogonal comparison for one factor by those for a given orthogonal comparison of the other factor, we will obtain an orthogonal comparison

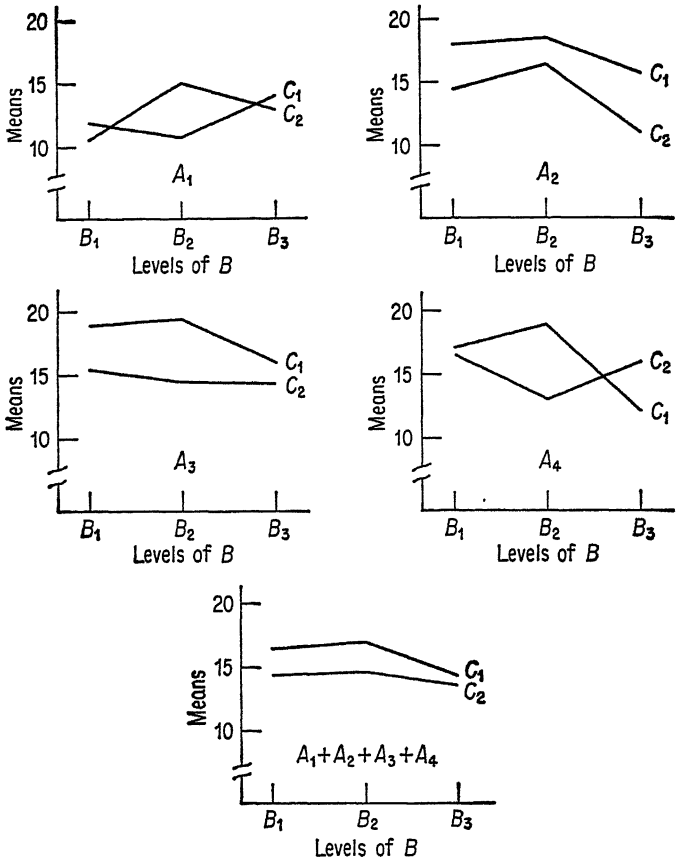


Figure 13.1 Means for levels of C at each level of B for A_1, A_2, A_3, A_4, and averaged over the levels of A. The $B \times C$ interaction averaged over the levels of A is nonsignificant. The fact that the $A \times B \times C$ interaction is significant means that the $B \times C$ interactions are not of the same form for the different levels of A. The nature of these interactions is shown for each level of A. Original data given in Table 13.6.

We again assume that the levels of the factors are fixed and do not represent random selections from any larger populations. Our conclusions, therefore, are to be restricted to the conditions we have actually investigated. Thus, the within-treatments sum of squares, when divided by the

corresponding to a component of the interaction sum of squares with 1 d.f. Thus, the interaction sum of squares with 4 d.f. can be analyzed into four component parts, each with 1 d.f. Some examples are given in the questions and problems at the end of the chapter. Since these comparisons are not apt to be planned comparisons, it is suggested that they be evaluated by means of Scheffé's test.

96 d.f. associated with it, provides us with an estimate of experimental error for testing the significance of the other mean squares.

The values of F shown in the analysis of variance table must be interpreted in terms of the number of degrees of freedom involved and these vary, depending upon the mean square being tested for significance. To simplify matters, we have starred the values of F which are significant with $\alpha = .05$. The interpretation of the significant values of F follows the same pattern as described in the previous chapter.

As a matter of interest, since it is the smallest mean square, we have graphed the $B \times C$ interaction to show that the lines for C_1 and C_2 plotted against the levels of B have much the same form. This graph is shown at the bottom of Figure 13.1. The means that are plotted were obtained from Table (5) of Table 13.6. The $A \times B \times C$ interaction is significant, indicating that the two-factor interactions are not the same for the levels of the third factor. For purposes of comparison, we have also graphed in Figure 13.1 the $B \times C$ interactions for each level of A. The means plotted in these graphs were obtained from the values given in Tables (1), (2), (3), and (4) of Table 13.6.

DIRECT CALCULATION OF A THREE-FACTOR INTERACTION

In some factorial experiments it will be necessary to calculate directly the sum of squares for a three-factor interaction. We will now illustrate a method of calculating the sum of squares for a three-factor interaction. The procedure is perfectly general and can be applied to obtain any interaction, regardless of the number of factors and the number of levels of the factors involved in the interaction.

The interaction sum of squares to be calculated is that of $A \times B \times C$. Examine the data of Table 13.1. The cell entries there are the sums of $n = 5$ observations for each treatment. The sums of Table 13.1 may be rearranged in the manner of Table 13.6. We have first separated the treatments according to the 4 levels of A and then according to the levels of B and C.

Consider only Table (1). For the data in this table we can calculate a sum of squares between the 6 cells. We can also calculate the sum of squares for rows (C) and for columns (B). The row sum of squares would be the sum of squares for C averaged over the levels of B with the level of A (A_1) held constant. The column sum of squares would be the sum of squares for B averaged over the levels of C with the level of A (A_1) held constant. If we subtract the row and column sums of squares from the sum of squares between cells, we shall have an interaction sum of squares. This procedure involves nothing new. We have used this method of calculation before in obtaining a two-factor interaction sum of squares. The interaction

Table 13.6 Tables for the Calculation of the $A \times B \times C$ Interaction Sum of Squares

Table (1)

	A_1			Σ
	B_1	B_2	B_3	
C_1	60	54	70	184
C_2	58	76	66	200
Σ	118	130	136	384

Table (2)

	A_2			Σ
	B_1	B_2	B_3	
C_1	90	92	76	258
C_2	72	82	56	210
Σ	162	174	132	468

Table (3)

	A_3			Σ
	B_1	B_2	B_3	
C_1	94	98	80	272
C_2	78	74	72	224
Σ	172	172	152	496

Table (4)

	A_4			Σ
	B_1	B_2	B_3	
C_1	86	96	60	242
C_2	84	64	78	226
Σ	170	160	138	468

Table (5)

	$A = A_1 + A_2 + A_3 + A_4$			Σ
	B_1	B_2	B_3	
C_1	330	340	286	956
C_2	292	296	272	860
Σ	622	636	558	1,816

obtained from Table (1), however, is the $B \times C$ interaction with the level of A held constant and we designate this interaction sum of squares by $A_1(B \times C)$.

The process described could be repeated for Tables (2), (3), and (4). We would thus have the interactions: $A_1(B \times C)$, $A_2(B \times C)$, $A_3(B \times C)$, and $A_4(B \times C)$. The necessary calculations for these interactions are as follows:

$$\text{Table (1): Between cells} = \frac{(60)^2}{5} + \frac{(58)^2}{5} + \cdots + \frac{(66)^2}{5} - \frac{(384)^2}{30} = 67.20$$

$$\text{Rows} = \frac{(184)^2}{15} + \frac{(200)^2}{15} - \frac{(384)^2}{30} = 8.53$$

$$\text{Columns} = \frac{(118)^2}{10} + \frac{(130)^2}{10} + \frac{(136)^2}{10} - \frac{(384)^2}{30} = 16.80$$

$$A_1(B \times C) = 67.20 - 8.53 - 16.80 = 41.87$$

Table (2): Between cells $= \frac{(90)^2}{5} + \frac{(72)^2}{5} + \cdots + \frac{(56)^2}{5} - \frac{(468)^2}{30} = 176\ 00$

Rows $= \frac{(258)^2}{15} + \frac{(210)^2}{15} - \frac{(468)^2}{30} = 76.80$

Columns $= \frac{(162)^2}{10} + \frac{(174)^2}{10} + \frac{(132)^2}{10} - \frac{(468)^2}{30} = 93.60$

$A_2(B \times C) = 176.00 - 76.80 - 93.60 = 5\ 60$

Table (3): Between cells $= \frac{(94)^2}{5} + \frac{(78)^2}{5} + \cdots + \frac{(72)^2}{5} - \frac{(496)^2}{30} = 116.27$

Rows $= \frac{(272)^2}{15} + \frac{(224)^2}{15} - \frac{(496)^2}{30} = 76.80$

Columns $= \frac{(172)^2}{10} + \frac{(172)^2}{10} + \frac{(152)^2}{10} - \frac{(496)^2}{30} = 26.67$

$A_3(B \times C) = 116.27 - 76.80 - 26.67 = 12.80$

Table (4): Between cells $= \frac{(86)^2}{5} + \frac{(84)^2}{5} + \cdots + \frac{(78)^2}{5} - \frac{(468)^2}{30} = 188.80$

Rows $= \frac{(242)^2}{15} + \frac{(226)^2}{15} - \frac{(468)^2}{30} = 8.53$

Columns $= \frac{(170)^2}{10} + \frac{(160)^2}{10} + \frac{(138)^2}{10} - \frac{(468)^2}{30} = 53.60$

$A_4(B \times C) = 188\ 80 - 8.53 - 53.60 = 126.67$

Summing the interactions of $B \times C$ for each level of A, we have

$$\sum A(B \times C) = 41\ 87 + 5.60 + 12.80 + 126\ 67 = 186\ 94$$

Now we have already calculated the $B \times C$ interaction averaged over the levels of A, that is, the $B \times C$ interaction for Table (5). Table (5), for example, is identical with Table 13.4, and for the $B \times C$ interaction we obtained a sum of squares equal to 12.60. Then, the sum of squares for the three-factor interaction, $A \times B \times C$ may be obtained by subtracting the $B \times C$ interaction sum of squares from the sum of the interactions of $B \times C$ for the separate levels of A. Thus,

$$A \times B \times C = \sum A(B \times C) - B \times C$$

Substituting in the above formula, we obtain

$$A \times B \times C = 186.94 - 12.60 = 174.34$$

which checks, within rounding errors, with the value previously found for the $A \times B \times C$ interaction sum of squares.

The method described above for calculating the sum of squares for a three-factor interaction can be varied to fit the needs of a particular factorial experiment. For example, the three-factor interaction sum of squares might have been obtained by calculating the $A \times C$ interactions for each level of B, summing, and then subtracting the $A \times C$ interaction obtained from the two-way table in which $A \times C$ entries are summed over the levels of C. Thus, in general, a three-factor interaction sum of squares will be given by

$$\begin{aligned} A \times B \times C &= \sum A(B \times C) - B \times C \\ &= \sum B(A \times C) - A \times C \\ &= \sum C(A \times B) - A \times B \end{aligned} \tag{13.2}$$

and a four-factor interaction sum of squares will be given by

$$\begin{aligned} A \times B \times C \times D &= \sum A(B \times C \times D) - B \times C \times D \\ &= \sum B(A \times C \times D) - A \times C \times D \\ &= \sum C(A \times B \times D) - A \times B \times D \\ &= \sum D(A \times B \times C) - A \times B \times C \end{aligned} \tag{13.3}$$

and a five-factor interaction sum of squares will be given by

$$\begin{aligned} A \times B \times C \times D \times E &= \sum A(B \times C \times D \times E) - B \times C \times D \times E \\ &= \sum B(A \times C \times D \times E) - A \times C \times D \times E \\ &= \sum C(A \times B \times D \times E) - A \times B \times D \times E \\ &= \sum D(A \times B \times C \times E) - A \times B \times C \times E \\ &= \sum E(A \times B \times C \times D) - A \times B \times C \times D \end{aligned} \tag{13.4}$$

A similar series of equations may be written for any interaction involving six factors and so on. A general proof of these equations is given by Edwards and Horst (1950).

Since a three-or-more-factor interaction may be found in a variety of ways, it is worth while to examine the data to determine the set of tables that will require the least effort as far as calculations are concerned. We have not, for example, taken the most economical set of tables in the present problem. The calculations would be reduced considerably if we had taken $A \times B \times C = \sum C(A \times B) - A \times B$ instead of taking $A \times B \times C = \sum A(B \times C) - B \times C$. The first equation would require only two tables, one for $C_1(A \times B)$ and one for $C_2(A \times B)$, whereas the second equation, as we have seen, requires four tables.

MANY FACTORS WITH MANY LEVELS

In the analysis of variance, a given set of observations consisting of one observation for each treatment is called a *replication.* Thus, if we have $k = 5$ treatments with $n = 10$ observations for each treatment, the experiment would be described as having 10 replications. It is often true in research that we are interested in many factors, each with many levels. It should be obvious, however, that with only 5 factors, each at 3 levels, one replication of the factorial experiment will require 243 observations. To be able to obtain an error mean square based on a within-treatments sum of squares requires at least one additional replication or at least 243 additional observations. If observations are associated with subjects, this would mean that this particular factorial experiment would require a total of 486 subjects for two replications. This number may exceed the available subjects. It is also possible that the time required to make 486 observations, if the subjects are available, may be excessive.

A Single Replication of a Factorial Experiment

Several solutions to the problem of a large number of treatments have been suggested. One solution is to have only one replication. In this instance, no estimate of experimental error corresponding to the mean square within treatments is available. When this is the case, the highest-order interaction or a combination of the higher-order interactions is used as an estimate of experimental error.[3] In the example cited, this would be either the five-factor interaction mean square with 32 d.f. or a pooled mean square based upon the four-factor and five-factor interactions. Since there are 5 four-factor interactions, each with 16 d.f., the latter mean square would have $(5)(16) + 32 = 112$ d.f.

The danger involved in using interactions as estimates of experimental error is that they may be of importance, that is, significant. When this is the case, the significance of the main effects and the other lower-order interactions will be underestimated.

Fractional Replication

Another solution to the problem of a large number of treatments is based upon the notion of *fractional replication.* For a 2^n factorial experiment, that is, with all factors at two levels, only a certain fraction, $\frac{1}{2}$ or $\frac{1}{4}$ or $\frac{1}{8}$, of the possible treatments are tested. Fractional replication is also possible with factorial experiments in which all factors are at three levels.

[3] The expectation is that these interactions will be negligible and also of little experimental interest.

Fractional replication is based upon the assumption that certain comparisons or effects, usually the interactions, are negligible or unimportant.

Consider, for example, the 2^3 factorial experiment of Table 12.7. We see there that the $A \times B \times C$ interaction is based upon a comparison of

$$(A_1B_1C_1 + A_1B_2C_2 + A_2B_1C_2 + A_2B_2C_1) - (A_1B_1C_2 + A_1B_2C_1 + A_2B_1C_1 + A_2B_2C_2)$$

Now assume that we make use of fractional replication by replicating only the first four treatments or the last four, it does not matter which. We thus sacrifice the information on the $A \times B \times C$ interaction provided by complete replication. If the first four treatments are replicated n times, then for the comparisons between the treatment sums we would have:

Comparison	$A_1B_1C_1$	$A_1B_2C_2$	$A_2B_1C_2$	$A_2B_2C_1$
A	1	1	−1	−1
B	1	−1	1	−1
C	1	−1	−1	1
$A \times B$	1	−1	−1	1
$A \times C$	1	−1	1	−1
$B \times C$	1	1	−1	−1

We observe, in this instance, that the sum of squares for A is identical with the $B \times C$ sum of squares, the B sum of squares is identical with the $A \times C$ sum of squares, and the C sum of squares is identical with the $A \times B$ sum of squares. If the various two-factor interactions are negligible, then ½ replication will provide estimates of the A effect, the B effect, and the C effect. Of course, for the 2^3 factorial we would ordinarily not use fractional replication, since the total number of treatments is only 8. If we add another factor with two levels, however, then the number of treatments will be 16. With still another factor at two levels, the number of treatment combinations will be 32. In this instance, fractional replication may be useful. It can be shown, for example, that for the 2^5 factorial experiment a ½ fractional replication will provide separate estimates of the main effects and the two-factor interactions, provided the higher-order interactions are negligible.

Details concerning the use of fractional replication for the 2^n series of factorial experiments can be found in Cochran and Cox (1957). Their discussion of the uses and of the hazards of fractional replication should be studied carefully by the experimenter who wishes seriously to consider the use of fractional replication. The discussions of fractional replication by Kempthorne (1952) and Cox (1958) are also of value.

FACTORIAL EXPERIMENTS WITH RANDOMIZED BLOCKS

We have discussed factorial experiments with respect to randomized groups designs A factorial experiment may also be used in connection with a randomized blocks design. Suppose, for example, we have a 2×2 factorial experiment and we have some reason to believe that subjects of comparable levels of intelligence will tend to be more homogeneous in their performance on the dependent variable, in the absence of treatment effects, than subjects selected at random. We have intelligence test scores for 20 subjects and on the basis of these scores the subjects are arranged into 5 blocks of 4 subjects each. Within each block the 4 treatments are assigned at random with one subject for each treatment.

Randomly assigning the treatments within each block, we may have the following arrangement:

	Randomization			
Block 1	A_1B_2	A_2B_1	A_1B_1	A_2B_2
Block 2	A_1B_1	A_2B_1	A_2B_2	A_1B_2
Block 3	A_2B_2	A_1B_2	A_2B_1	A_1B_1
Block 4	A_2B_2	A_1B_2	A_1B_1	A_2B_1
Block 5	A_1B_2	A_1B_1	A_2B_1	A_2B_2

The actual observations obtained in the experiment may then be rearranged in the manner of Table 13.7.

Table 13.7 A 2×2 Factorial Experiment in a Randomized Blocks Design

Block	Treatments				Σ
	A_1B_1	A_1B_2	A_2B_1	A_2B_2	
1	5	4	4	7	20
2	4	5	3	5	17
3	3	6	2	6	17
4	2	3	1	4	10
5	1	2	0	3	6
Σ	15	20	10	25	70

	B_1	B_2	Σ
A_1	15	20	35
A_2	10	25	35
Σ	25	45	70

Error Mean Square of a Randomized Blocks Design

When we discussed the randomized blocks design previously, we referred to the error mean square as the residual mean square It is perhaps now clear that the residual mean square is an interaction mean square. For the randomized blocks design, the total sum of squares corresponds to a between-cells sum of squares. The block sum of squares corresponds to a row sum of squares and the treatment sum of squares corresponds to a column sum of squares. In the randomized blocks design, we subtracted the block (row) and treatment (column) sums of squares from the total (between-cells) to obtain the residual. Formula (13.1) shows that the residual sum of squares of the randomized blocks design is identical with the rows $\times$ columns (blocks $\times$ treatments) sum of squares. Each block of the randomized blocks design constitutes one *replication* of the experiment. Thus, for the randomized blocks design, we have the following identity

$$\begin{aligned} \text{Error} = \text{Residual} &= \text{Blocks} \times \text{treatments} \\ &= \text{Replications} \times \text{treatments} \end{aligned} \tag{13.5}$$

The error sum of squares of Table 13.8 is a pooled sum of squares based upon the following interactions:

$$\begin{aligned} \text{Error} = (\text{Replications} \times A) &+ (\text{Replications} \times B) \\ &+ (\text{Replications} \times A \times B) \end{aligned}$$

Each of the three sums of squares on the right has 4 d.f. and each divided by its degrees of freedom is assumed to be an estimate of a common error variance. By pooling the sums of squares and the associated degrees of freedom, we obtain the error mean square of Table 13.8 with 12 d.f. This error mean square is used in testing the significance of the A, B, and $A \times B$ mean squares. The error mean square can be obtained most easily by subtracting the treatment and block sums of squares from the total sum of squares.

ORGANISMIC VARIABLES AS FACTORS

We now consider experiments in which one of the factors of interest corresponds to an *organismic* variable. As we have pointed out earlier, organismic variables refer to various ways in which we may classify subjects. As examples, we have sex, intelligence, attitude, and various other ways in which individuals can be said to differ.

In our discussion of randomized blocks designs, we pointed to the possible use of organismic variables as a basis for arranging subjects into blocks. By placing together in the same block subjects who are homogeneous with respect to some characteristic, we hope to obtain a smaller estimate

of experimental error than we would with a randomized groups design. Thus, with the randomized blocks design, we use the organismic variable in an attempt to reduce the estimate of experimental error. The organismic variable itself is not of experimental interest

The types of experiments with which we are now concerned resemble factorial experiments in which one of the factors is an organismic variable and this factor is of experimental interest. In the factorial experiments we have discussed previously, the various factors all referred to treatments, a given treatment consisting of one level from each factor. Treatments were then assigned at random to subjects and our tests of significance were concerned only with the treatment effects. Thus, when the levels of a factor represent treatment differences, these are effectively randomized over the subjects involved in the experiment. On the other hand, when the levels of a factor correspond to differences between subjects, there is no way in which the experimenter can randomly assign the levels to the subjects.

Suppose, for example, we are interested in performance, under a standardized condition, of subjects classified as anxious and subjects classified as nonanxious. Thus we might say that this experiment is concerned with the factor of anxiety and that we have two levels of the factor. But it should be obvious that the levels of anxiety are associated with subjects and there is no way in which they can be randomly assigned to the subjects, as can be done with the levels of a factor corresponding to treatment differences. An organismic factor thus differs in a very important way from a treatment factor. If the factor represents a treatment, then the levels of the factor can be randomly assigned to subjects. If the factor represents an organismic variable, then the levels cannot be randomly assigned to the subjects. The anxiety level of a subject, in other words, is a property of the subject and not something that can be randomly assigned to him.

When treatments are randomly assigned to subjects or subjects are randomly assigned to treatments, we anticipate that the process of randomization will randomize individual differences between the treatment groups. Thus, if we obtain a significant difference between the treatment means, we can interpret this difference as being produced by the differences in the treatments themselves. Suppose, for example, that the treatments consist of one and two presentations of a passage. If the treatments are randomly assigned to the subjects and if we find a significant difference in retention between the two treatment groups, we have a basis for concluding that the difference is the result of the treatments themselves and not the result of systematic differences between the subjects in the two treatment groups.

Suppose now that we test a group of anxious and a group of nonanxious subjects under the *same* treatment and obtain for each subject some measure of learning. Assume we find that the means of the two groups differ, with the anxious group having a lower mean than the nonanxious group. It is

of importance to emphasize again that randomization is not involved in this study. We cannot, therefore, say that the difference in level of anxiety of the two groups produces the difference in the means We have established that there is a relationship between level of anxiety and learning, but the finding that two variables are related does not in and of itself imply anything about which variable is cause and which is effect. Before we could conclude that the differences in anxiety produce the differences in learning, we would have to be able to demonstrate that level of anxiety was the *only* way in which the two groups differed. But to be able to demonstrate that the two groups differ only with respect to level of anxiety would be exceedingly difficult. If we were to find that differences in level of anxiety are, in turn, correlated or associated with differences in intelligence, for example, then it would be just as logical or illogical to attribute the difference in the learning means to differences in intelligence as to differences in anxiety.

That we cannot regard a factor corresponding to an organismic variable in the same manner as a factor corresponding to treatment differences is important. But this difference in the nature of the factors does not rule out the possibility of incorporating organismic factors in an experiment in a meaningful way. It is to this problem that we now turn.

AN EXPERIMENT WITH AN ORGANISMIC FACTOR

Suppose that an organismic factor of interest is anxiety. We designate this factor by A and we have two levels, A_1 representing a "high" level of anxiety and A_2 representing a "low" level of anxiety. We have 20 subjects at each level, obtained by administering a test of anxiety to a large group of males and then selecting the 20 with the highest scores on the test and the 20 with the lowest scores. We also have a treatment factor B with two levels, B_1 and B_2. Let B_1 correspond to punishment administered for each wrong response and B_2 a reward for each correct response, with the dependent variable being a measure of learning. *Within* each level of anxiety, we *randomly* assign $n = 10$ subjects to each level of B. For each anxiety level considered separately, the experimental design is thus a randomized groups design.

Let us assume that the pooled sum of squares, based upon the variation within each of the 4 groups, is equal to 900.0. This pooled sum of squares will have $k(n - 1) = 4(10 - 1) = 36$ d.f. The sums for each of the 4 groups are given in Table 13.9. Then the sum of squares between groups will be equal to

$$\text{Between groups} = \frac{(120)^2}{10} + \frac{(180)^2}{10} + \cdots + \frac{(100)^2}{10} - \frac{(530)^2}{40} = 347.5$$

Table 13.9 Sums for a 2 × 2 Factorial Experiment Where *A* Is an Organismic Factor—Each Cell Entry Is the Sum of $n = 10$ Observations

	B_1	B_2	$\sum$
A_1	120	180	300
A_2	130	100	230
$\sum$	250	280	530

and this sum of squares can be further analyzed into the sums of squares for A, B, and $A \times B$. Thus

$$A = \frac{(300)^2}{20} + \frac{(230)^2}{20} - \frac{(530)^2}{40} = 122.5$$

$$B = \frac{(250)^2}{20} + \frac{(280)^2}{20} - \frac{(530)^2}{40} = 22.5$$

$$A \times B = \frac{[(120 + 100) - (180 + 130)]^2}{(4)(10)} = 202.5$$

Table 13.10 Analysis of Variance for the 2 × 2 Factorial Experiment of Table 13.9

Source of Variation	Sum of Squares	d.f	Mean Square	*F*
A	122.5	1	122.5	4.9
B	22.5	1	22.5	
$A \times B$	202 5	1	202 5	8.1
Within treatments	900.0	36	25.0	
Total	1,247 5	39		

The results of our analysis are summarized in Table 13.10. The A mean square is significant, indicating that the means for the high anxiety subjects and the low anxiety subjects differ significantly. This comparison, however, is of little experimental interest since, in the absence of randomization of the A levels, we do not have a clear interpretation of the comparison. The high and low anxiety subjects, for example, may differ in a variety of other respects as well as in level of anxiety.

If the B mean square had been significant, this would show that the means for B_1 and B_2, averaged over the levels of anxiety, differed significantly. Since randomization for B_1 and B_2 occurred within each level of anxiety, no difficulty would be involved in interpreting a significant B effect. For the same reason we have no difficulty in interpreting the significant $A \times B$ interaction mean square, since, with randomization within anxiety levels, each high anxiety subject had an equal chance of being

assigned to B_1 or B_2 and each low anxiety subject had an equal chance of being assigned to B_1 or B_2. Note, for example, that formula (12.1) which gives the $A \times B$ comparison is based upon the difference $(A_1B_1 + A_2B_2) - (A_1B_2 + A_2B_1)$. Thus the sum $A_1B_1 + A_2B_2$ is based upon 10 randomly selected high anxiety subjects and 10 randomly selected low anxiety subjects. The sum $A_1B_2 + A_2B_1$ is also based upon 10 randomly selected high anxiety subjects and 10 randomly selected low anxiety subjects. The $A \times B$ comparison thus meets the requirements of randomness and can be interpreted in the same manner as we interpret a treatment effect based upon random assignment.

The significance of the $A \times B$ interaction mean square shows that the difference between B_1 and B_2 for A_1 (high anxiety subjects) is not of the same form as the difference between B_1 and B_2 for A_2 (low anxiety subjects). For the high anxiety group, the mean for B_1 is $120/10 = 12.0$ and the mean for B_2 is $180/10 = 18.0$, whereas, for the low anxiety group, the two means are $130/10 = 13.0$ and $100/10 = 10.0$, respectively. Of course, we are still not in a position where we can attribute this finding to anxiety level without other supporting evidence that this is the case. Our argument may be helped if we had predicted from theoretical considerations that high anxiety and low anxiety subjects should respond differentially to the B treatments in the manner actually observed.

In general, in experiments involving an organismic factor, the difference between the levels of the organismic factor is of little experimental importance. It is primarily the presence or absence of interaction between the organismic factor and the treatment factor that is of interest. The nature of the interaction, if one is found, can be examined by means of the graphic methods described previously.

In the present experiment, we may wish to test for the significance of the difference between B_1 and B_2 separately for each level of A. It can easily be shown that these two tests represent orthogonal comparisons, although they are not orthogonal with the $A \times B$ interaction comparison. It is suggested, therefore, if these and additional comparisons are to be made, that they be tested by means of Scheffé's test, described earlier.

QUESTIONS AND PROBLEMS

1. An experiment involves factor A, which is varied in 3 ways, factor B, which is varied in 2 ways, factor C, which is varied in 2 ways, and factor D, which is varied in 3 ways. The experiment is replicated with $n = 5$ subjects for each treatment combination. (*a*) Set up the summary analysis of variance table showing the sources of variation and the number of degrees of freedom associated with each. (*b*) How would you calculate the $A \times B \times D$ interaction sum of squares?

2. A factorial experiment involves two factors, A and B, with A varied in 4 ways and B varied in 3 ways. The treatment combinations are replicated with $n = 5$ observations for each. Results are given below:

A_1			A_2			A_3			A_4		
B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3
38	54	65	24	21	35	36	35	35	45	45	34
45	34	86	43	67	45	81	36	65	55	98	65
22	54	62	56	98	76	22	54	67	34	65	65
23	23	26	75	46	89	23	65	76	34	34	43
45	32	42	43	55	98	45	78	55	45	54	36

Use the analysis of variance to analyze the results of the experiment.

3. Child (1946) designed an experiment to test the hypothesis that preference for a more distant goal object, when found, is the result of experience in previous situations. "The experiment was planned so that if this assumption was correct, certain influences of previous learning would be exhibited" (p. 3). The factors introduced were as follows: the sex of the children used as subjects in the experiment; the sex of the experimenter present during the test situation; the nature of the barrier introduced between the subject and the distant goal object; and the type of instructions given to the child. "The basic technique of these experiments was to place children in the position of having to choose between two desirable goals, one of which was more accessible than the other, and to observe their reactions" (p. 5).

Subjects were school children in grades 1 through 7. They were divided into groups of 34 to 45 subjects each. The data given are in terms of the percentage choosing the more distant goal. Child states that the percentages are "close enough to 50, to suggest an adequate approximation to the assumption of normal distribution of sampling errors" (pp. 18–19). The analysis of variance was applied, however, making use of the inverse sine transformation. The values of F obtained with the transformation were slightly different, but no conclusions concerning significance were changed by the analysis of the data on the transformed scale. The results are given below:

	Male Subjects		Female Subjects	
	Cued Instructions	Noncued Instructions	Cued Instructions	Noncued Instructions
Male experimenter				
Table barrier	43	36	13	21
Ladder barrier	40	50	24	32
Female experimenter				
Table barrier	33	41	39	30
Ladder barrier	55	46	37	43

(*a*) Compute the various sums of squares. You may find the procedure of setting up a table with orthogonal coefficients a convenient method of calculation. (*b*) Note that in this experiment there is no mean square within treatments and that for tests of significance the higher-order interactions must be used for an

error mean square. For tests of significance, Child used the pooled sum of squares for all interactions with 11 d.f. What are some of the problems and assumptions involved in this procedure? (*c*) Note also that sex of the subjects corresponds to an organismic variable. If this mean square is significant, what interpretation may be made? (*d*) Sex of the experimenter corresponds to a treatment factor in which the levels (male and female) may be randomly assigned to subjects. As in the other factorial experiments we have discussed, however, levels of this factor do not represent a random sampling from a larger population of male and female experimenters. Thus, if significant, the conclusions should be restricted to the particular female and male experimenter involved in the experiment and not generalized beyond the two actually used.

4. We have the following results for a factorial experiment in which we have only one replication:

$A_1B_1C_1 = 40$	$A_2B_1C_1 = 30$
$A_1B_1C_2 = 60$	$A_2B_1C_2 = 60$
$A_1B_1C_3 = 70$	$A_2B_1C_3 = 60$
$A_1B_2C_1 = 60$	$A_2B_2C_1 = 60$
$A_1B_2C_2 = 20$	$A_2B_2C_2 = 10$
$A_1B_2C_3 = 20$	$A_2B_2C_3 = 60$
$A_1B_3C_1 = 50$	$A_2B_3C_1 = 20$
$A_1B_3C_2 = 90$	$A_2B_3C_2 = 90$
$A_1B_3C_3 = 50$	$A_2B_3C_3 = 10$

Find the various mean squares. Note that three of the interaction mean squares are fairly large relative to the mean squares for the main effects. It is entirely possible that one more of these interactions would be significant if we had available a mean square based upon replication which could be used in the test of significance.

5. We have a factorial experiment in which A is varied in two ways and B is varied in three ways. Results are given below:

A_1			A_2		
B_1	B_2	B_3	B_1	B_2	B_3
19	43	53	30	49	64
10	42	51	24	43	61
21	41	57	25	49	68
15	44	57	30	53	56
20	42	68	28	44	60
24	49	60	34	46	55
16	46	48	31	46	54
22	39	47	32	56	68
18	48	60	27	54	59
18	39	60	32	53	57

(*a*) Analyze the data using the analysis of variance. (*b*) Examine the $A \times B$ interaction, regardless of whether or not it is significant, by graphic methods.

6. Define, briefly: (*a*) replication, (*b*) fractional replication.

7. If a significant three-factor interaction is obtained in an experiment, how could one go about examining the nature of the interaction?

8. Describe an experiment in which one might expect to find a significant three-factor interaction. Explain why you would expect this result.

9. Suppose we have a factorial experiment with a levels of A and b levels of B. Then it is always possible to analyze the sum of squares for A into a set of $a - 1$ mutually orthogonal comparisons, each with 1 d.f. Furthermore, it is possible to analyze the sum of squares for B into a set of $b - 1$ mutually orthogonal comparisons. Consider a simple example, with two levels of A and three levels of B. The sum for each treatment combination is given below:

Comparison	A_1			A_2		
	B_1	B_2	B_3	B_1	B_2	B_3
	10	15	20	15	10	30
1	1	1	1	−1	−1	−1
2	2	−1	−1	2	−1	−1
3	0	1	−1	0	1	−1
4	2	−1	−1	−2	1	1
5	0	1	−1	0	−1	1

(*a*) Assuming that $n = 10$ subjects have been randomly assigned to each treatment combination, find the sum of squares between groups with 5 d.f. (*b*) Analyze the sum of squares between groups into the comparisons: A, B, and $A \times B$. (*c*) Now examine the comparisons shown in the above table. Are they mutually orthogonal? (*d*) Each of the comparisons given in the table will have 1 d.f. Note that the sum of the sums of squares for comparisons (2) and (3) is equal to the sum of squares for B. The sum of squares for B with 2 d.f., in other words, has been analyzed into the two orthogonal comparisons shown, each with 1 d.f. (*e*) Note that the coefficients for comparison (4) are obtained by multiplying the coefficients of (1) and (2). Similarly, the coefficients for comparison (5) are obtained by multiplying the coefficients of (1) and (3). Furthermore, the sum of the sums of squares for the comparisons (4) and (5) is equal to the $A \times B$ interaction sum of squares. We have, in other words, analyzed the $A \times B$ interaction sum of squares with 2 d.f. into the two orthogonal comparisons shown in (4) and (5) each with 1 d.f.

10. As another example, assume that the T(hirst), H(unger), and S(ex) drives of rats are each at three levels or intensities, 1, 2, and 3.

Comparison	Thirst			Hunger			Sex		
	1	2	3	1	2	3	1	2	3
	10	18	28	12	20	30	10	12	15
1	1	1	1	1	1	1	−2	−2	−2
2	1	1	1	−1	−1	−1	0	0	0
3	−1	−1	2	−1	−1	2	−1	−1	2
4	−1	1	0	−1	1	0	−1	1	0
5	−1	−1	2	−1	−1	2	2	2	−4
6	−1	1	0	−1	1	0	2	−2	0
7	−1	−1	2	1	1	−2	0	0	0
8	−1	1	0	1	−1	0	0	0	0

(*a*) Assuming $n = 10$ rats have been randomly assigned to each treatment combination, find the sum of squares between treatment groups with 8 d.f. (*b*) Analyze the sum of squares between treatment groups into the comparisons: drive, intensity, and drive $\times$ intensity. (*c*) Now examine the comparisons shown in the above table. Are they mutually orthogonal? (*d*) Each of the comparisons given in the table will have 1 d.f. Into what comparisons has the sum of squares for drive been analyzed? Into what comparisons has the sum of squares for intensity been analyzed? Into what comparisons has the sum of squares for drive $\times$ intensity been analyzed?

14

TREND ANALYSIS

INTRODUCTION

In studies of learning, our interest is centered in improvement or change in performance as a result of practice. In a sense, practice can be considered a factor with the successive periods of practice or trials as levels. If, for example, we were interested in change in performance over 5 trials, and if 50 subjects were available, we might randomly assign the 5 levels in such a way that we have 10 subjects for each level or amount of practice. One group would be given but a single trial, another two trials, a third three trials, and so on. For each subject in each group, we would use only the final measure, under the assumption that it provides an estimate of performance for a specified number of trials. The analysis of the experimental data would be the same as for a randomized groups design.

If we used the measure obtained on Trial 1 for the subjects, it would also be possible to arrange the subjects in blocks of 5 such that within each block the subjects are relatively homogeneous with respect to performance on Trial 1. Then using random methods, one subject in each block would be assigned to each of the 5 levels or trials. In this instance, the analysis of variance would correspond to that of a randomized blocks design.

However, most experimenters would feel that both of the above procedures are inefficient in the sense that 4 observations, all but the last, would be discarded for the subjects with 5 trials. Similarly, we would discard the first three observations for those subjects receiving 4 trials, the first two observations for those subjects receiving 3 trials, and the first observation for those subjects receiving 2 trials. The argument would be that the same amount of information could be obtained by giving a single group of 10 subjects 5 trials. Thus, each subject would have a score or measure for each trial. In this instance, *each subject would correspond to a block* of 5 subjects in the randomized blocks design. We should note, however, that the trials (treatments) would not be randomized within each block, as they would be in a randomized blocks design, but rather would occur in exactly the same sequence or order for each block.

In this chapter we shall discuss the analysis of experiments concerned with the trend of a series of means in which more than one observation or

measurement is made on each subject. All of the experiments to be described will involve the notion that a single subject corresponds to what we have called a block in our discussion of randomized blocks designs In these experiments, the primary objective is to study the *trend* of the means over the successive trials. The observations for each trial are obtained under a *standard* condition and it is assumed that any differences found between the trial means are the result of the differing amounts of practice. An extension of this experimental design involves the introduction of one or more factors. These factors may be treatments which can be randomized or organismic factors which cannot be randomized.

An examination of the means for a series of trials may reveal that the trend is either upward or downward, that is, the means may either increase with successive trials or they may decrease. Now such a trend can, of course, occur as a result of random variation. From the experimenter's point of view, the important question is whether the upward or downward trend can be regarded as meeting the requirements of statistical significance or whether it should be regarded as a random or chance affair. Similarly, the trend of the means, in addition to being downward or upward, as the case may be, may also show a bend or degree of curvature. Again, if there is a bend or curvature in the trend, we wish to be able to determine whether the curvature is such as to meet the requirements of statistical significance.

If trial means are available for two or more treatment groups, then we may wish to determine whether there are significant differences between certain characteristics of the trends of means for the various treatment groups. For example, if we have trial means for two different treatment groups we may have reason to believe that the trend of the means for one treatment should be sharply downward whereas the trend for the other treatment should be only slightly downward. An examination of the trial means for each treatment group may indicate that the trends are in accord with expectation. A corresponding test of significance provides a basis for determining whether the difference in the trends for the two treatments is significant.

It should be emphasized that the methods of analysis described in this chapter are not concerned with the problem of finding an equation that will describe the trend of the trial means. The problem of curve fitting, of course, is one of importance. Our concern in this chapter, however, is in providing the experimenter with methods for determining whether certain characteristics of the trend of the trial means for a single group are statistically significant or whether they can be attributed to random variation.

It may also be emphasized that before undertaking the analyses and tests of significance described, it is always advisable to plot the means for the successive trials for each treatment or experimental condition. Examination of these plots prior to the data analysis is of value in that the plots

suggest what the data analysis may confirm. Furthermore, it is advisable to present either the means or the plots in reporting the data analysis since they assist others in understanding the results of the analysis.

TRIAL MEANS: ONE STANDARD CONDITION

Table 14.1 gives measures of performance on each of 5 subjects for each of 3 trials. We find the total sum of squares, the sum of squares for

Table 14.1 Observations Obtained for 5 Subjects on 3 Trials

Subjects	Trials			Σ
	1	2	3	
1	3	7	10	20
2	7	9	11	27
3	2	4	7	13
4	2	6	10	18
5	6	9	12	27
Σ	20	35	50	105

subjects (blocks or rows) and the sum of squares for trials (columns) in the usual way Thus

$$\text{Total} = (3)^2 + (7)^2 + \cdots + (12)^2 - \frac{(105)^2}{15} = 144.00$$

$$\text{Trials} = \frac{(20)^2}{5} + \frac{(35)^2}{5} + \frac{(50)^2}{5} - \frac{(105)^2}{15} = 90.00$$

$$\text{Subjects} = \frac{(20)^2}{3} + \frac{(27)^2}{3} + \cdots + \frac{(27)^2}{3} - \frac{(105)^2}{15} = 48.67$$

It is evident, from formula (13.1), that if we subtract the trial and subject sums of squares from the total, we shall have the subjects $\times$ trials interaction sum of squares. Thus

$$S\text{'s} \times \text{trials} = \text{Total} - \text{subjects} - \text{trials} \tag{14.1}$$

or, for the present example,

$$S\text{'s} \times \text{trials} = 144.00 - 48.67 - 90.00 = 5.33$$

Table 14.2 summarizes the analysis of variance. The fact that the S's $\times$ trials sum of squares is not very large indicates that the form of the

Table 14.2 Analysis of Variance of the Data of Table 14.1

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Trials	90 00	2	45.00	67.2
Subjects	48 67	4	12.17	
S's × trials	5 33	8	.67	
Total	144 00	14		

learning curve for the various subjects is much the same. This could be examined graphically by plotting each subject's learning curve for the 3 trials Using the S's × trials mean square as our estimate of experimental error, we find that the trial mean square is highly significant.

The sum of squares for trials, with 2 d.f., may be partitioned into a sum of squares for linear regression, the linear component of the trend, with 1 d.f and a sum of squares for curvature, the quadratic component of the trend, also with 1 d.f. From Table XI, with $k = 3$ trials, we have as the orthogonal coefficients for the linear component, -1, 0, and 1. Then the sum of squares for the linear component of the trend of the trial means, as given by formula (10.8), is

$$\text{Linear component} = \frac{[(-1)(20) + (0)(35) + (1)(50)]^2}{(5)(2)} = 90.0$$

and we note that this sum of squares is exactly equal to the sum of squares for trials. Thus the sum of squares for curvature, the quadratic component, will be zero. The trend of the trial means can thus be accurately represented by a straight line. All of this, of course, is obvious in this simple example, and could easily be shown by plotting the trial means. In actual experiments involving actual data, results are seldom so obvious or so simple.

TRIAL MEANS: DIFFERENT TREATMENTS

In learning experiments we may be interested not only in the change of performance over a series of trials under a standard condition, but under different experimental treatments. Various other factors, in addition to practice, may influence the shape of the learning curve. For example, we may be interested in the progress of learning under different dosages of a drug or under different drugs at a standard dosage. Frequency of reinforcement may be varied in several ways and we may wish to know whether the different levels of reinforcement influence the shape of the learning curve. For treatment factors such as those described, randomization in the assignment of the levels of the factor is possible and should be used.

Randomization

Let us suppose we are interested in the influence of three drugs, each at a standard dosage, on learning. We designate this treatment factor as A and let the three drugs be represented by A_1, A_2, and A_3. We have 15 subjects available and the drugs are assigned at random in such a way that we have $n = 5$ subjects for each drug. Let us assume also that we have decided to test each subject on three trials, that is, we shall have 3 observations for each subject. We designate the trials by B and the succeeding trials by B_1, B_2, and B_3. The layout of the experiment, with the levels of A randomized is as follows for 15 subjects:

$A_1B_1B_2B_3$	$A_3B_1B_2B_3$	$A_1B_1B_2B_3$
$A_2B_1B_2B_3$	$A_2B_1B_2B_3$	$A_3B_1B_2B_3$
$A_1B_1B_2B_3$	$A_3B_1B_2B_3$	$A_2B_1B_2B_3$
$A_3B_1B_2B_3$	$A_1B_1B_2B_3$	$A_2B_1B_2B_3$
$A_3B_1B_2B_3$	$A_2B_1B_2B_3$	$A_1B_1B_2B_3$

Each subject corresponds to a block and the levels of A have been randomized over the blocks.[1]

Sums of Squares

The actual observations made for each subject can be rearranged in the manner of Table 14.3. We find the total sum of squares, the sum of squares for subjects (blocks or rows), and the sum of squares for trials (columns) in the usual manner. Thus

$$\text{Total} = (2)^2 + (2)^2 + \cdots + (10)^2 - \frac{(340)^2}{45} = 369.11$$

$$\text{Subjects} = \frac{(13)^2}{3} + \frac{(18)^2}{3} + \cdots + \frac{(27)^2}{3} - \frac{(340)^2}{45} = 175.78$$

$$\text{Trials} = \frac{(80)^2}{15} + \frac{(110)^2}{15} + \frac{(150)^2}{15} - \frac{(340)^2}{45} = 164.44$$

[1] Other than the fact that the levels of B have not been randomized, this design is similar to a *split-plot* design. In the split-plot design the levels of A are randomized over blocks and within each block the levels of B are randomized. The split-plot design in psychological research will be discussed later. For further discussion of this design, see Snedecor (1956), Cochran and Cox (1957), or Kempthorne (1952).

Table 14.3 Observations for 3 Groups with Each Group Tested under a Different Drug and with 3 Trials for Each Subject

Drugs	Subjects	Trials			Σ
		B_1	B_2	B_3	
	1	2	4	7	13
	2	2	6	10	18
A_1	3	3	7	10	20
	4	7	9	11	27
	5	6	9	12	27
	1	5	6	10	21
	2	4	5	10	19
A_2	3	7	8	11	26
	4	8	9	11	28
	5	11	12	13	36
	1	3	4	7	14
	2	3	6	9	18
A_3	3	4	7	9	20
	4	8	8	10	26
	5	7	10	10	27
	Σ	80	110	150	340

If we subtract the subject and trial sums of squares from the total, we will have the S's $\times$ trials interaction sum of squares. Thus

$$S\text{'s} \times \text{trials} = 369.11 - 175.78 - 164.44 = 28.89$$

The number of subjects tested with each drug is $n = 5$. We have $a = 3$ drugs and $b = 3$ trials. Then the total sum of squares will have $nab - 1 = (5)(3)(3) - 1 = 44$ d.f. The sum of squares for subjects will have $na - 1 = (5)(3) - 1 = 14$ d.f. and the sum of squares for trials will have $b - 1 = 2$ d.f. The S's $\times$ trials sum of squares will have $(na - 1)(b - 1) = (14)(2) = 28$ d.f.

The sum of squares between subjects with 14 d.f. can be analyzed into two component parts. One of these sums of squares will be the sum of squares for drugs with $a - 1 = 2$ d f. The sum of squares for drugs can be obtained from the marginal entries of Table 14.4. Thus

$$\text{Drugs} = \frac{(105)^2}{15} + \frac{(130)^2}{15} + \frac{(105)^2}{15} - \frac{(340)^2}{45} = 27.78$$

If we subtract the sum of squares for drugs from the sum of squares for subjects, we will have a residual equal to $175.78 - 27.78 = 148.00$ with

$14 - 2 = 12$ d.f. This residual sum of squares is the pooled sum of squares between subjects in each drug group. Thus

$$S\text{'s in } A_1 = \frac{(13)^2}{3} + \frac{(18)^2}{3} + \cdots + \frac{(27)^2}{3} - \frac{(105)^2}{15} = 48.67$$

$$S\text{'s in } A_2 = \frac{(21)^2}{3} + \frac{(19)^2}{3} + \cdots + \frac{(36)^2}{3} - \frac{(130)^2}{15} = 59.33$$

$$S\text{'s in } A_3 = \frac{(14)^2}{3} + \frac{(18)^2}{3} + \cdots + \frac{(27)^2}{3} - \frac{(105)^2}{15} = 40.00$$

The sum of these three sums of squares is equal to $48.67 + 59.33 + 40.00 = 148.00$ and this is exactly equal to the value we obtained above by subtraction. Each of the separate sums of squares has $n - 1 = 4$ d.f. and since we have $a = 3$ of these sums of squares, the pooled sum of squares has $a(n - 1) = 3(5 - 1) = 12$ d.f.

The S's $\times$ trials interaction sum of squares with 28 d.f. can be analyzed into two component parts. One of these sums of squares will be the drugs $\times$

Table 14.4 Sums for Drugs and Trials for the Data of Table 14.3

Drugs	Trials			Σ
	B_1	B_2	B_3	
A_1	20	35	50	105
A_2	35	40	55	130
A_3	25	35	45	105
Σ	80	110	150	340

trials interaction sum of squares with $(a - 1)(b - 1) = (3 - 1)(3 - 1) = 4$ d.f. This sum of squares can be obtained by first calculating the sum of squares between the cells of Table 14.4. Thus

$$\text{Between cells} = \frac{(20)^2}{5} + \frac{(35)^2}{5} + \cdots + \frac{(45)^2}{5} - \frac{(340)^2}{45} = 201.11$$

We have already calculated the row or drug (A) sum of squares and the column or trial (B) sum of squares for Table 14.4. Then, by subtraction, we obtain

$$\text{Drugs} \times \text{trials} = 201.11 - 27.78 - 164.44 = 8.89$$

If we subtract the drugs $\times$ trials sum of squares from the subjects $\times$ trials sum of squares we obtain a residual which is equal to $28.89 - 8.89 = 20.00$, and this residual will have $28 - 4 = 24$ d.f. This residual sum of squares is

the pooled S's $\times$ trials interaction calculated separately for each drug. For each drug, for example, we could obtain the S's $\times$ trials interaction sum of squares and each of these sums of squares would have $(n - 1)(b - 1) = 8$ d.f. Since we have $a = 3$ of these interactions, the pooled interaction will have $(3)(8) = 24$ d.f.

Summary of the Analysis

Table 14.5 summarizes the calculations. This analysis of variance summary table differs from those we have presented previously in that we

Table 14.5 Analysis of Variance of the Data of Table 14.3

Source of Variation		Sum of Squares	d.f.	Mean Square	F
A:	Drugs	27.78	2	13.89	1.13
	Error (a)	148.00	12	12.33	
B:	Trials	164.44	2	82.22	99.06
$A \times B$:	Drugs $\times$ trials	8.89	4	2.22	2.67
	Error (b)	20.00	24	.83	
	Total	369.11	44		

have in the table two mean squares designated as error mean squares rather than a single error mean square. The error mean square designated (a) is based upon the pooled sum of squares between subjects and is the appropriate error mean square for testing the significance of the A or drug effect. The error mean square designated (b) is based upon the pooled subjects $\times$ trials interactions and is the appropriate error mean square for testing the significance of the B or trial effect and the $A \times B$ or drugs $\times$ trials interaction.

For the A effect, we have $F = 13.89/12.33 = 1.13$ with 2 and 12 d.f., and this is not a significant value. Since the three means, A_1, A_2, and A_3, have been averaged over the 3 trials, they correspond to a general over-all measure of performance for each drug. If the A mean square had been significant, we would have concluded that the A means differed significantly. Its nonsignificance shows that the average performance over 3 trials is not significantly different for the different drugs.

For the B effect, we have $F = 82.22/83 = 99.06$ with 2 and 24 d.f. and this is a highly significant value. The 3 trial means, B_1, B_2, and B_3, have been averaged over the 3 drugs and we conclude that these means differ significantly.

Testing the $A \times B$ interaction mean square for significance, we have $F = 2.22/.83 = 2.67$ with 4 and 24 d.f. The tabled value of F for 4 and 24 d.f., at the 5 per cent level, is 2.78 and our obtained value just misses

being declared significant at this level. If the $A \times B$ or drugs $\times$ trials mean square had been significant, this would have meant that the learning curves for the three drugs were not of the same form. To examine the interaction, we plot curves for each drug in Figure 14.1. If these curves were all exactly parallel, the interaction sum of squares would be zero. As it is, there is some tendency for the curves to have somewhat different forms, but, with $\alpha = .05$, we cannot say that the forms differ significantly.

In the next section, we shall illustrate some additional analyses that can be made with respect to the trend of the over-all trial means and of the differences in the trends of the trial means for the separate drug groups.

In experiments of the kind described, where the levels of the A factor are randomized over blocks (subjects), we will, in general, find that the error mean square designated (a) is greater than the error mean square designated (b) in Table 14.5. Our primary interest is usually in the B (trials) effect and the $A \times B$ interaction, and both of these are tested for significance with the error mean square designated (b). The significance or lack of significance of the $A \times B$ interaction tells us whether or not the trend of the trial means is of the same form for the various levels of A. Since the levels of A have been randomized over the subjects, a significant $A \times B$ interaction should not occur as a result of systematic organismic differences between the subjects in the various A groups.

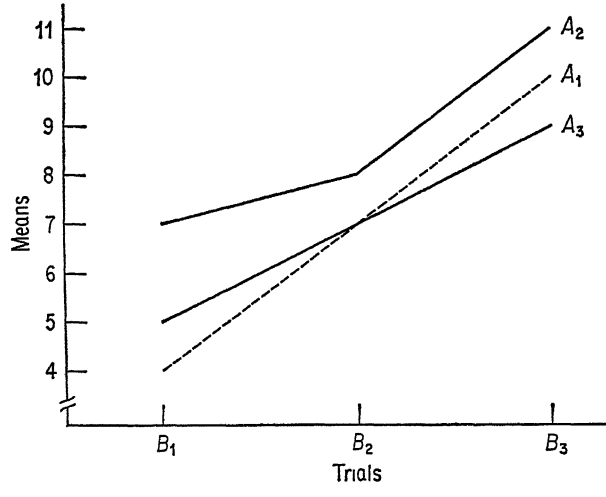


Figure 14.1 Means for levels of *A* at each level of *B*. Levels of *A* correspond to different drugs. Original data given in Table 14.4.

TRIAL MEANS: A TREATMENT FACTOR AND AN ORGANISMIC FACTOR

Using data from an experiment by Grant and Patel (1957), Grant (1956) has provided an example of trend analysis in which two factors, each at two or more levels, were involved as well as trials. An organismic factor of interest was anxiety. On the basis of a test of anxiety, two groups, a group of 12 high anxious and a group of 12 low anxious subjects, were selected. We designate the anxiety factor by A and the two levels by A_1 and A_2, with A_1 corresponding to the high anxious group and A_2 to the

Table 14.6 Perseverative Error Scores at Different Stages (C) for Anxiety Groups (A) at Three Levels of Shock (B)*

Anxiety-Shock Combination	Subjects	Stages					Σ
		C_1	C_2	C_3	C_4	C_5	
A_1B_1	1	1.4	1.0	.0	1.4	1.0	4.8
A_1B_1	2	1.7	2.4	2.0	3.7	2.2	12.0
A_1B_1	3	1.7	4.1	1.7	4.0	3.2	14.7
A_1B_1	4	2.8	1.0	2.4	3.7	3.2	13.1
A_1B_2	1	2.0	1.7	2.4	1.0	1.4	8.5
A_1B_2	2	1.7	1.4	.0	.0	.0	3.1
A_1B_2	3	1.4	.0	1.0	.0	.0	2.4
A_1B_2	4	1.4	1.4	1.0	1.0	.0	4.8
A_1B_3	1	1.4	1.0	2.0	2.4	1.0	7.8
A_1B_3	2	1.4	2.4	1.0	1.4	2.8	9.0
A_1B_3	3	1.0	2.0	1.4	1.0	1.4	6.8
A_1B_3	4	4.2	1.7	2.2	4.9	4.7	17.7
A_2B_1	1	1.7	1.0	1.7	1.0	1.0	6.4
A_2B_1	2	3.0	1.7	1.0	1.0	1.0	7.7
A_2B_1	3	2.4	1.4	3.0	2.4	.0	9.2
A_2B_1	4	1.4	1.0	1.0	1.0	.0	4.4
A_2B_2	1	2.0	2.4	1.0	.0	.0	5.4
A_2B_2	2	2.8	1.0	1.0	1.0	1.0	6.8
A_2B_2	3	5.3	2.0	.0	.0	1.0	8 3
A_2B_2	4	2.0	4.1	1.4	1.4	1.4	10.3
A_2B_3	1	2.2	1.7	1.7	1.4	1.0	8.0
A_2B_3	2	4.4	.0	1.7	1.0	.0	7.1
A_2B_3	3	2.0	4.7	1.4	1.0	1.0	10.1
A_2B_3	4	2.8	1.7	2.0	.0	.0	6.5
Σ		54.1	42.8	34.0	35.7	28.3	194.9

* From Grant (1956).

low anxious group. The treatment factor in the experiment was shock and this factor had three levels. We let the shock factor be B and let B_1, B_2, and B_3 correspond to the three levels. The dependent variable consisted of perseverative error scores at 5 different stages of the Wisconsin Card Sorting Test. The various stages can be represented by C and the successive stages by C_1, C_2, C_3, C_4, and C_5.

The experimental design was a randomized blocks design, with each subject corresponding to a block or row. The individual observations given

Table 14.7 Sums for Anxiety (A), Shock (B), and Stages (C). Each Cell Entry Is the Sum of 4 Observations

		C_1	C_2	C_3	C_4	C_5	$\sum$
High: A_1	B_1	7.6	8 5	6.1	12 8	9.6	44.6
	B_2	6 5	4.5	4.4	2.0	1.4	18.8
	B_3	8.0	7.1	6 6	9 7	9.9	41.3
	$\sum$	22.1	20 1	17.1	24.5	20.9	104.7
Low: A_2	B_1	8.5	5.1	6.7	5.4	2 0	27.7
	B_2	12.1	9 5	3.4	2.4	3.4	30.8
	B_3	11 4	8.1	6.8	3.4	2 0	31.7
	$\sum$	32.0	22 7	16 9	11.2	7.4	90.2

in Table 14.6 are based upon a square root transformation of the original error scores. The cell entries of Table 14.7 are the sums of $n = 4$ observations corresponding to each ABC combination.

Sums of Squares

We find the total sum of squares, the sum of squares between subjects, and the sum of squares for stages in the usual manner. Thus

$$\text{Total} = (1.4)^2 + (1.7)^2 + \cdots + (0)^2 - \frac{(194.9)^2}{120} = 165.9799$$

$$\text{Subjects} = \frac{(4.8)^2}{5} + \frac{(12.0)^2}{5} + \cdots + \frac{(6.5)^2}{5} - \frac{(194.9)^2}{120} = 59.4639$$

$$\text{Stages} = \frac{(54.1)^2}{24} + \frac{(42.8)^2}{24} + \cdots + \frac{(28.3)^2}{24} - \frac{(194.9)^2}{120} = 16.3678$$

Subtracting the sum of squares for stages and the sum of squares for subjects from the total, we obtain

$$S\text{'s} \times \text{stages} = 165.9799 - 59.4639 - 16.3678 = 90.1482$$

The degrees of freedom for the sums of squares we have calculated can be obtained in the usual way. Thus the total sum of squares will have $120 - 1 = 119$ d.f., the sum of squares for subjects will have $24 - 1 = 23$ d.f., the sum of squares for stages will have $5 - 1 = 4$ d.f., and the S's $\times$ stages sum of squares will have $(24 - 1)(5 - 1) = 92$ d.f.

Partitioning the Subject Sum of Squares

The sum of squares between subjects with 23 d.f. can be analyzed into the sum of squares for anxiety with 1 d.f., the sum of squares for shock with 2 d.f., the sum of squares for anxiety $\times$ shock with 2 d.f., and the pooled sum of squares between subjects for each combination of anxiety and shock with $(2)(3)(4 - 1) = 18$ d.f.

Table 14.8 Two-Way Table for Anxiety and Shock. Each Cell Entry Is the Sum of 20 Observations

Anxiety	Shock			$\sum$
	B_1	B_2	B_3	
A_1	44 6	18.8	41.3	104 7
A_2	27 7	30 8	31 7	90 2
$\sum$	72 3	49 6	73.0	194 9

The sums of squares for anxiety (A) and shock (B) and the anxiety $\times$ shock ($A \times B$) interaction sum of squares can be obtained from Table 14.8. Thus

$$\text{Anxiety} = \frac{(104.7)^2}{60} + \frac{(90.2)^2}{60} - \frac{(194.9)^2}{120} = 1.7521$$

$$\text{Shock} = \frac{(72.3)^2}{40} + \frac{(49.6)^2}{40} + \frac{(73.0)^2}{40} - \frac{(194.9)^2}{120} = 8.8611$$

$$\text{Between cells} = \frac{(44.6)^2}{20} + \frac{(27.7)^2}{20} + \cdots + \frac{(31.7)^2}{20} - \frac{(194.9)^2}{120} = 21.9054$$

$$\text{Anxiety} \times \text{shock} = 21.9054 - 1.7521 - 8.8611 = 11.2922$$

The pooled sum of squares between subjects in the various groups can be obtained by direct calculation, in the manner described previously, but it can also be obtained by subtracting the sum of squares for anxiety, shock, and anxiety $\times$ shock, from the sum of squares between subjects with the grouping ignored. Thus

$$\text{Pooled between } S\text{'s} = 59.4639 - 1.7521 - 8.8611 - 11.2922 = 37.5585$$

and this sum of squares has been designated as error (a) in Table 14.11.

Partitioning the *S*'s × Stages Sum of Squares

The *S*'s × stages sum of squares can also be analyzed in the manner

Table 14.9 Two-Way Table for Anxiety and Stages. Each Cell Entry Is the Sum of 12 Observations

Anxiety	Stages					Σ
	C_1	C_2	C_3	C_4	C_5	
A_1	22.1	20.1	17.1	24.5	20.9	104.7
A_2	32.0	22 7	16.9	11 2	7 4	90.2
Σ	54 1	42.8	34.0	35 7	28.3	194.9

described previously. To obtain the anxiety × stages sum of squares, we first find the sum of squares between the cells of Table 14.9, and then subtract the sum of squares for anxiety and the sum of squares for stages, which we have already calculated, from the sum of squares between cells. Thus

$$\text{Between cells} = \frac{(22.1)^2}{12} + \frac{(32.0)^2}{12} + \cdots + \frac{(7.4)^2}{12} - \frac{(194.9)^2}{120} = 35.6991$$

$$\text{Anxiety} \times \text{stages} = 35.6991 - 1.7521 - 16.3678 = 17.5792$$

The anxiety × stages sum of squares will have $(2 - 1)(5 - 1) = 4$ d.f.

Similarly, to find the shock × stages sum of squares, we first calculate the sum of squares between cells of Table 14.10. Then we subtract two sums

Table 14.10 Two-Way Table for Shock and Stages. Each Cell Entry Is the Sum of 8 Observations

Shock	Stages					Σ
	C_1	C_2	C_3	C_4	C_5	
B_1	16.1	13.6	12.8	18 2	11.6	72.3
B_2	18.6	14.0	7.8	4.4	4.8	49.6
B_3	19.4	15.2	13.4	13.1	11.9	73.0
Σ	54.1	42.8	34.0	35.7	28.3	194.9

of squares we have already calculated, the sum of squares for shock and the sum of squares for stages, from the sum of squares between cells. Thus

$$\text{Between cells} = \frac{(16.1)^2}{8} + \frac{(18.6)^2}{8} + \cdots + \frac{(11.9)^2}{8} - \frac{(194.9)^2}{120} = 35.8483$$

$$\text{Shock} \times \text{stages} = 35.8483 - 8.8611 - 16.3678 = 10.6194$$

The shock × stages sum of squares will have $(3 - 1)(5 - 1) = 8$ d.f.

The anxiety × shock × stages interaction sum of squares can be obtained by subtraction. Consider, for example, Table 14.7 where we show the sums for each anxiety-shock group for each stage. The sum of squares between the cells of this table is equal to

$$\text{Between cells} = \frac{(7.6)^2}{4} + \frac{(6.5)^2}{4} + \cdots + \frac{(2.0)^2}{4} - \frac{(194.9)^2}{120} = 73.5324$$

with 29 d.f. The row sum of squares for this table has 5 d.f. and is equal to the sum of the sums of squares for anxiety, shock, and anxiety × shock. The column sum of squares, with 4 d.f., for the table is equal to the sum of squares for stages. The sum of squares between the cells of Table 14.7 is equal to the sum of the sums of squares for anxiety (A), shock (B), anxiety × shock ($A \times B$), stages (C), anxiety × stages ($A \times C$), shock × stages ($B \times C$), and anxiety × shock × stages ($A \times B \times C$). Since we have calculated all of these sums of squares except the last one, it can be obtained by subtraction. Thus the anxiety × shock × stages ($A \times B \times C$) interaction sum of squares will be given by

$$\begin{aligned} A \times B \times C &= \text{Between cells} - A - B - C - A \times B - A \times C - B \times C \\ &= 73.5324 - 1.7521 - 8\,8611 - 16.3678 \\ &\qquad - 11.2922 - 17.5792 - 10.6194 \\ &= 7.0606 \end{aligned}$$

The anxiety × stages sum of squares (17.5792), the shock × stages sum of squares (10.6194), and the anxiety × shock × stages sum of squares (7.0606) are all part of the S's × stages sum of squares (90.1482). Subtracting the sums of squares we have just calculated from the S's × stages sum of squares, we obtain the *pooled* S's × stages sum of squares. Thus

$$\text{Pooled } S\text{'s} \times \text{stages} = 90.1482 - 17.5792 - 10.6194 - 7.0606 = 54.8890$$

The pooled S's × stages sum of squares could, of course, be calculated directly by finding the S's × stages sum of squares separately for each anxiety-shock combination. For a single anxiety-shock group, the S's × stages sum of squares would have $(4 - 1)(5 - 1) = 12$ d.f. Since we have 6 different anxiety-shock groups, the pooled S's × stages sum of squares will have $(6)(12) = 72$ d.f. This pooled sum of squares has been designated as error (b) in Table 14.11.

Summary of the Analysis

Table 14.11 summarizes the analysis of variance. As in the previous example, we have two error mean squares. The one designated (a) is the appropriate error mean square for testing the significance of the A effect,

the B effect, and the $A \times B$ interaction. Error (b) is the appropriate error mean square for the other tests. We note, in this example, as in the previous one, error (b) is smaller than error (a), and this will usually be the case.

Table 14.11 Analysis of Variance of the Data of Table 14.6

Source of Variation		Sum of Squares	d f.	Mean Square	F
A:	Anxiety	1.7521	1	1.7521	
B:	Shock	8.8611	2	4.4306	2.12
$A \times B$.	Anxiety × shock	11.2922	2	5.6461	2 71
	Error (a)	37.5585	18	2.0866	
C.	Stages	16.3678	4	4 0920	5.37*
$A \times C$:	Anxiety × stages	17.5792	4	4.3948	5.77*
$B \times C$:	Shock × stages	10 6194	8	1 3274	1.74
$A \times B \times C$:	Anxiety × shock × stages	7 0606	8	.8823	1.16
	Error (b)	54 8890	72	.7623	
	Total	165.9799	119		

Only the two values of F marked with an asterisk are significant. The over-all stage means, averaged over anxiety and shock, differ significantly. The significant anxiety × stages mean square tells us that the trend of the stage means does not have the same form for the two anxiety groups. In Figure 14.2 the stage means, averaged over the levels of shock, are shown for each anxiety group. We observe that the trend for the low anxiety group (A_2) is downward and appears to be approximately linear. For the high

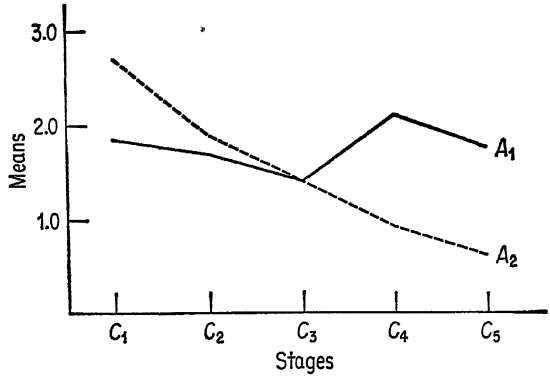


Figure 14.2 Means for levels of A at each level of C. A_1 and A_2 correspond to high and low anxiety, respectively. Original data given in Table 14.9.

anxiety group (A_1) the trend is downward for the first three stages but then it tends slightly upward.

TREND ANALYSIS OF THE OVER-ALL STAGE SUMS

Linear Component of the Trend

The trend of the over-all stage means is shown in Figure 14.3. The sum of squares for the over-all trend, stages, is equal to 16.3678, with 4 d.f. To obtain the sum of squares for the linear component of the trend, we make use of the orthogonal coefficients of Table XI, in the Appendix. With 5 stages, the orthogonal coefficients are -2, -1, 0, 1, and 2. Multiplying each of the over-all stage sums by these coefficients we have

$$(-2)(54.1) + (-1)(42.8) + (0)(34.0) + (1)(35.7) + (2)(28.3) = -58.7$$

We have $n = 24$ observations for each sum, and the sum of the squared orthogonal coefficients is 10. Then, by formula (10.8), we have

$$\text{Linear component} = \frac{(-58\,7)^2}{(24)(10)} = 14.3570$$

as the sum of squares for the linear component of the over-all trend with 1 d.f.

To test for the significance of the linear component of the over-all trend, we use error mean square (b) of Table 14.11. Thus, we have $F = 14.3570/.7623 = 18.83$ with 1 and 72 d.f. This is a highly significant value and we conclude that the over-all stage means do show a linear trend. The direction of this trend is downward or negative as shown by the fact that

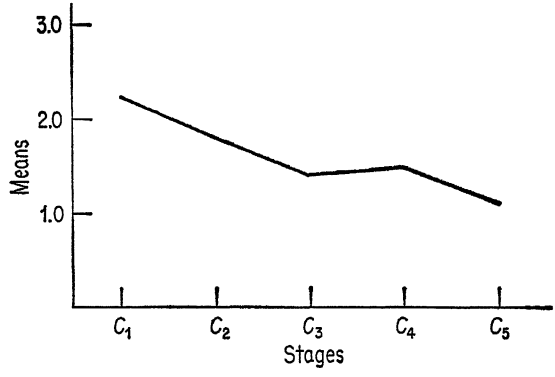


Figure 14.3 Means, averaged over levels of anxiety and shock, for each of 5 stages. Original data given in Table 14.6.

the numerator of the linear component, -58.7, is negative in sign. When the numerator of the linear component is positive in sign, this indicates an upward trend of the means.

Quadratic Component of the Trend

To determine whether there is a significant curvature in the trend of the over-all stage means, we use the orthogonal coefficients for the quadratic component. These coefficients, obtained from Table XI, are 2, -1, -2, -1, and 2 and the sum of the squared coefficients is 14. Multiplying the over-all stage sums by the coefficients, we have

$$(2)(54.1) + (-1)(42.8) + (-2)(34.0) + (-1)(35.7) + (2)(28.3) = 18.3$$

Then, the sum of squares for the quadratic component will be

$$\text{Quadratic component} = \frac{(18.3)^2}{(24)(14)} = .9967$$

with 1 d.f. To test the significance of this component, we have $F = .9967/.7623 = 1.31$ with 1 and 72 d.f. and this is not a significant value. We conclude that there is no significant curvature in the over-all trend of the stage means.

LINEAR COMPONENTS OF INTERACTIONS WITH STAGES

Groups × Stages Interaction

Let $k = 6$, the number of anxiety-shock groups. The stage sums for the 6 groups are given in Table 14.12. The groups × stages interaction sum of squares for this table is equal to 35.2592 and has $(k - 1)(c - 1) = 20$ d.f. The mean square is equal to $35.2592/20 = 1.7630$ and, testing for significance, we have $F = 1.7630/.7623 = 2.31$, a significant value for 20 and 72 d.f. We did not make this test in Table 14.11, since we analyzed the groups × stages sum of squares into its three component parts: anxiety × stages with 4 d.f., shock × stages with 8 d.f., and anxiety × shock × stages with 8 d.f. These three sums of squares, as can be seen in Table 14.11, are 17.5792, 10.5960, and 7.0840, respectively, and the sum of the sums of squares is equal to 35.2592, the sum of squares for groups × stages.

It should be clear that the groups × stages mean square has to do with the trend of the stage means for each of the 6 anxiety-shock groups. The fact that the groups × stages mean square is significant indicates that the form of the trend of the stage means for the different groups is not the same.

We wish now to examine further the trend of the stage means for the various groups. We can find the sum of squares for the linear and quadratic components of the groups × stages sum of squares. We analyzed the groups × stages sum of squares in the analysis of variance of Table 14.11 into the three component parts: anxiety × stages, shock × stages, and anxiety × shock × stages. So also we can analyze the linear component of the groups × stages sum of squares into three linear components: anxiety × stages, shock × stages, and anxiety × shock × stages. In the same way, we can partition the quadratic component of the groups × stages sum of squares into the quadratic components for anxiety × stages, shock × stages, and anxiety × shock × stages.

Multiplying each of the row entries of Table 14.12 by the orthogonal coefficients for the linear component, we obtain the entries in column D_1.

Table 14.12 Trend Analysis Table for the Stage Sums of Table 14.7

		Orthogonal Coefficients					Comparisons	
Linear:		−2	−1	0	1	2	Linear	Quadratic
Quadratic:		2	−1	−2	−1	2		
		C_1	C_2	C_3	C_4	C_5	D_1	D_2
	B_1	7.6	8.5	6.1	12.8	9.6	8 3	.9
A_1	B_2	6.5	4.5	4.4	2.0	1.4	−12.7	.5
	B_3	8.0	7.1	6.6	9.7	9.9	6.4	5.8
	B_1	8.5	5.1	6.7	5.4	2.0	−12.7	−2.9
A_2	B_2	12.1	9.5	3.4	2.4	3.4	−24.5	12.3
	B_3	11.4	8.1	6.8	3.4	2.0	−23.5	1.7
	$\sum$	54.1	42.8	34.0	35.7	28.3	−58 7	18 3

For example, the first entry is

$$(-2)(7.6) + (-1)(8.5) + (0)(6.1) + (1)(12.8) + (2)(9.6) = 8.3$$

The last entry in this column, −58.7, is obtained by multiplying the over-all stage sums by the orthogonal coefficients and this must be equal to $\sum D_1$. We thus have a check on the accuracy of the calculations. The sum of the squared orthogonal coefficients is $\sum a_{.1}{}^2 = 10$ and for each group we have $n = 5$ subjects. Then the linear component of the groups × stages interaction sum of squares will be given by a general formula. Thus

$$\text{Linear component of interaction} = \frac{\sum_1^k D_1{}^2}{n \sum a_{.1}{}^2} - \frac{\left(\sum_1^k D_1\right)^2}{kn \sum a_{.1}{}^2} \qquad (14.2)$$

The first term on the right of formula (14.2) is the sum of $k = 6$ sums of squares of linear components, one for each group, and each of these has 1 d.f. The second term on the right is the linear component for the over-all trend and also has 1 d.f. Thus, the sum of squares of formula (14.2) will have $k - 1$ d.f. or, in the present case, $6 - 1 = 5$ d.f. Formula (14.2) provides a measure of the differences between the k comparisons, linear components, given by the first term on the right. Since it can be shown, in the present instance, that the sum of squares of formula (14.2) is the linear component of an interaction sum of squares, groups $\times$ stages, we have designated it as such.

Substituting in formula (14.2) with the appropriate values from Table 14.12, we have

$$\frac{(8.3)^2 + (-12.7)^2 + \cdots + (-23.5)^2}{(4)(10)} - \frac{(-58.7)^2}{(6)(4)(10)} = 25.2662$$

as the linear component of the groups $\times$ stages sum of squares and this sum of squares has 5 d.f. The mean square provides a measure of the differences between the linear components of the group trends.

Anxiety $\times$ Stages

Let us now obtain the linear component of the anxiety $\times$ stages sum of squares. To obtain D_1 for the high anxiety group, we would multiply the stage sums for this group by the orthogonal coefficients. It should be obvious that the results of this multiplication would give us the same value as $(8.3) + (-12.7) + (6.4) = 2.0$ Similarly, if we multiply the stage sums for the low anxiety group by the orthogonal coefficients we would obtain the same value as $(-12.7) + (-24.5) + (-23.5) = -60.7$. We have $n = 12$ subjects in each of the $k = 2$ anxiety groups. Then, substituting in formula (14.2), we have as the linear component of the anxiety $\times$ stages interaction sum of squares

$$\frac{(2.0)^2 + (-60.7)^2}{(12)(10)} - \frac{(-58.7)^2}{(2)(12)(10)} = 16.3804$$

and, for the reasons given previously, this component will have $k - 1$ or 1 d.f. The mean square provides a measure of the difference between the linear components of the trends of the means for the two anxiety groups.

Shock $\times$ Stages

We now find the linear component of the shock $\times$ stages sum of squares. To find D_1 for each shock group, we would multiply the stage sums for each group by the orthogonal coefficients. We have actually already performed this multiplication, in part, in Table 14.12. For example, D_1 for

the first level of shock will be equal to $(8.3) + (-12.7) = -4.4$; D_1 for the second level of shock will be equal to $(-12.7) + (-24\,5) = -37.2$; and D_1 for the third level of shock will be $(6\,4) + (-23.5) = -17.1$. We have $n = 8$ subjects in each of the $k = 3$ shock groups. Then substituting in formula (14.2), we have as the linear component of the shock $\times$ stages sum of squares

$$\frac{(-4.4)^2 + (-37\,2)^2 + (-17.1)^2}{(8)(10)} - \frac{(-58.7)^2}{(3)(8)(10)} = 6.8381$$

with $k - 1 = 2$ d.f. The mean square provides a measure of the differences between the linear components of the trends of the means of the three shock groups.

Anxiety $\times$ Shock $\times$ Stages

The linear component of the three-factor interaction, anxiety $\times$ shock $\times$ stages, can be obtained by subtraction. We subtract the linear components of anxiety $\times$ stages and shock $\times$ stages from the linear component of groups $\times$ stages to obtain the linear component of the three-factor interaction. Thus, remembering that we are subtracting linear components to obtain a linear component, we have

$$\begin{aligned}\text{Anxiety} \times \text{shock} \times \text{stages} &= \text{Groups} \times \text{stages} - \text{anxiety} \times \text{stages} \\ &\quad - \text{shock} \times \text{stages} \\ &= 25.2662 - 16.3804 - 6\,8381 \\ &= 2.0477\end{aligned}$$

The degrees of freedom for the linear component of the three-factor interaction can be obtained by substituting the degrees of freedom associated with each of the terms on the right in the above expression and subtracting. We have 5 d.f. for the groups $\times$ stages component, 1 d.f. for the anxiety $\times$ stages component, and 2 d.f for the shock $\times$ stages component. We thus have $5 - 1 - 2 = 2$ d.f. for the linear component of the three-factor interaction.

QUADRATIC COMPONENTS OF INTERACTIONS WITH STAGES

Multiplying each of the row entries in Table 14.12 by the orthogonal coefficients for the quadratic component, we obtain the entries in column D_2. The sum of the squared coefficients is $\sum a_{.2}{}^2 = 14$. Then, we have

$$\text{Quadratic component of interaction} = \frac{\sum_1^k D_2{}^2}{n \sum a_{.2}{}^2} - \frac{\left(\sum_1^k D_2\right)^2}{kn \sum a_{.2}{}^2} \quad (14.3)$$

This sum of squares, like that of formula (14.2), provides a measure of the differences between the k comparisons, quadratic components, given by the first term on the right. Since it can be shown, in the present example, that the sum of squares of formula (14.3) is the quadratic component of an interaction sum of squares, groups $\times$ stages, we have designated it as such.

The calculation of the quadratic components for the interactions of interest are summarized below. The calculations are based upon the same procedures we followed in calculating the linear components. Thus, we obtain the following quadratic components:

$$\text{Groups} \times \text{stages} = \frac{(.9)^2 + (.5)^2 + \cdots + (1.7)^2}{(4)(14)} - \frac{(18.3)^2}{(6)(4)(14)}$$

$$= \frac{197.29}{56} - \frac{334.88}{336}$$

$$= 3.5230 - .9967$$

$$= 2.5263$$

$$\text{Anxiety} \times \text{stages} = \frac{(.9 + .5 + 5.8)^2 + (-2.9 + 12.3 + 1.7)^2}{(12)(14)} - \frac{(18.3)^2}{(2)(12)(14)}$$

$$= \frac{(7.2)^2 + (11.1)^2}{168} - \frac{(18.3)^2}{336}$$

$$= \frac{175.05}{168} - \frac{334.88}{336}$$

$$= 1.0420 - .9967$$

$$= .0453$$

$$\text{Shock} \times \text{stages} = \frac{(.9 - 2.9)^2 + (.5 + 12.3)^2 + (5.8 + 1.7)^2}{(8)(14)} - \frac{(18.3)^2}{(3)(8)(14)}$$

$$= \frac{(-2.0)^2 + (12.8)^2 + (7.5)^2}{112} - \frac{(18.3)^2}{336}$$

$$= \frac{224.09}{112} - \frac{334.88}{336}$$

$$= 2.0008 - .9967$$

$$= 1.0041$$

Then, by subtraction, we obtain the quadratic component of the three-factor interaction. Thus

$$\text{Anxiety} \times \text{shock} \times \text{stages} = 2.5263 - .0453 - 1.0041$$

$$= 1.4769$$

Each of the quadratic components will have the same number of degrees of freedom as the corresponding linear components.

TESTS OF SIGNIFICANCE OF LINEAR AND QUADRATIC COMPONENTS

Table 14.13 summarizes the analysis. The error mean square (*b*) is obtained from Table 14.11.[2] The significance of the anxiety × stages mean square with $F = 21.49$ confirms our impression that the linear component of the trend of the stage means for the low anxiety group and the linear component of the trend for the high anxiety group differ significantly. We note, for example, that the linear component for the low anxiety group is $(-60.7)^2/(12)(10) = 30.7041$ whereas for the high anxiety group the linear component is $(2.0)^2/(12)(10) = .0333$. It is the marked discrepancy be-

Table 14.13 Analysis of Variance Showing the Linear and Quadratic Components of the Interactions with Stages Sums of Squares of Table 14.11

Source of Variation	Sum of Squares	d.f.	Mean Square	*F*
Linear components:				
Anxiety	16.3804	1	16.3804	21.49
Shock	6.8381	2	3.4190	4.49
Anxiety × shock	2.0477	2	1.0238	1.34
Groups	25 2662	5		
Quadratic components:				
Anxiety	.0453	1	.0453	
Shock	1.0041	2	.5020	
Anxiety × shock	1.4769	2	.7384	
Groups	2.5263	5		
Error (*b*)*	54.8890	72	.7623	

* From Table 14.11.

tween these two components that results in a significant, $F = 21.49$, anxiety × stages mean square.[3] Suppose, for example, that the linear components for the two anxiety groups were the same. Since $\sum D_1$ must

[2] In his analysis of this experiment, Grant (1956) obtained the linear and quadratic components of the error (*b*) sum of squares. The linear component of the error sum of squares, divided by its degrees of freedom, was then used to test the linear components of Table 14.13 for significance. Similarly, the quadratic component of the error sum of squares, divided by its degrees of freedom, was used to test the quadratic components of Table 14.13 for significance. We shall discuss the subdivision of the error sum of squares in the next section and show how to obtain linear and quadratic components.

[3] What we have called a significant difference between the linear components of the trends is also referred to, by some experimenters, as a significant difference between the *slopes* of the trends.

equal -58.7, then D_1 for each anxiety group would have to be equal to $-58.7/2 = -29\,35$. Then, for each group we would have as the linear component $(-29.35)^2/(12)(10) = 7.1785$. The linear component for the over-all trend, both anxiety groups combined, is $(-58.7)^2/(2)(12)(10) = 14.3570$. By substitution in formula (14.2) we see that, under this condition,

$$\text{Linear component of anxiety} \times \text{stages} = 7.1785 + 7.1785 - 14.3570 = 0$$

In the present analysis, we also find that the linear components of the trends for the separate shock groups differ significantly The graph of the stage means for the three shock groups is shown in Figure 14.4. For the three levels of shock, we have D_1 equal to -4.4, -37.2, and -17.1, respectively, and it is the failure of these to be comparable that results in the significant value of $F = 4.49$. Suppose, for example, that the D_1 values had been the same for each shock group. Since $\sum D_1$ must equal -58.7, each of the D_1 values would have to be equal to $-58.7/3 = -19.5667$. Then the linear components for each shock group would be equal to $(-19.5667)^2/(8)(10) = 4.7857$. For the over-all trend, all shock groups combined, we have as the linear component $(-58.7)^2/(3)(8)(10) = 14.3570$. Then, by substitution in formula (14.2), we see that under this condition

$$\text{Linear component of shock} \times \text{stages} = 4.7857 + 4.7857 + 4.7857 - 14.3570 = 0001$$

or zero except for rounding errors.

If the mean square for anxiety $\times$ shock of Table 14.13 had been significant, we would qualify our interpretations of the significant anxiety and significant shock mean squares accordingly. A significant anxiety $\times$

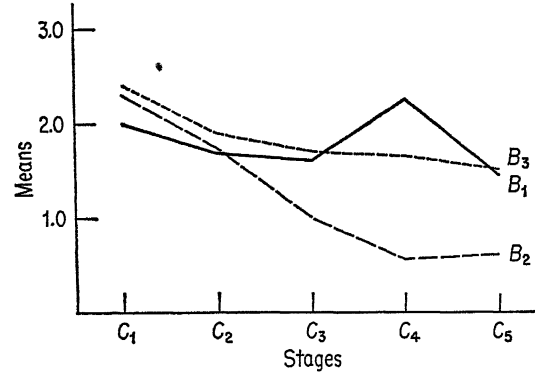


Figure 14.4 Means for levels of *B* at each level of *C*. Levels of *B* correspond to different levels of shock. Original data given in Table 14.10.

shock mean square, for example, would indicate that the difference found between the linear components of the stage means for the two anxiety groups is not independent of the level of shock

None of the mean squares for the quadratic comparisons given in Table 14.13 is significant, indicating that the difference in curvature of the trends for the two anxiety groups and the differences in curvature of the trends for the three shock groups are not significant.

TREND ANALYSIS OF THE DRUG EXPERIMENT

We now go back to the drug experiment presented earlier in the chapter and apply the kind of analysis we have just discussed. Table 14 14 repeats the sums for each drug group for each trial At the top of the table we show the orthogonal coefficients for the linear comparison and for the quadratic comparison. These coefficients were obtained from Table XI in the Appendix. We multiply each cell sum by the linear coefficients to obtain the entries in column D_1. The entries in column D_2 are obtained by multiplying each cell sum by the quadratic coefficients. The last entries in columns D_1 and D_2 are obtained by multiplying the over-all trial sums by

Table 14.14 Trend Analysis Table for the Trial Sums of Table 14.4

	Orthogonal Coefficients			Comparison	
Linear	−1	0	1		
Quadratic	1	−2	1	Linear	Quadratic
	B_1	B_2	B_3	D_1	D_2
A_1	20	35	50	30	0
A_2	35	40	55	20	10
A_3	25	35	45	20	0
$\sum$	80	110	150	70	10

the corresponding orthogonal coefficients and these values must be equal to $\sum D_1$ and $\sum D_2$, respectively.

Over-all Trial Means

We consider first the over-all trial means. The sum of the squared coefficients for the linear component of the trend is $\sum a_{.1}{}^2 = 2$, and the sum of the squared coefficients for the quadratic component is $\sum a_{.2}{}^2 = 6$. Then, for the linear component of the trend, we have, by substitution in formula (10.8),

$$\text{Linear component} = \frac{(70)^2}{(15)(2)} = 163.33$$

where 15 is the number of observations for each trial mean. This sum of squares has 1 d.f.

To test the linear component for significance we use error mean square (*b*) of Table 14.5. Thus, we have $F = 163.33/.83 = 196.78$, a highly significant value for 1 and 24 d.f.

To test for curvature of the over-all trend, we first find the quadratic component which is equal to

$$\text{Quadratic component} = \frac{(10)^2}{(15)(6)} = 1.11$$

with 1 d.f. Then we have $F = 1.11/.83 = 1.34$ and this is a nonsignificant value. The trend of the over-all trial means is essentially linear and there is no significant curvature. We note also that the sum of the two sums of squares, 163.33 and 1.11, each with 1 d.f., is equal to 164.44 or the sum of squares for trials with 2 d.f. Thus, we have made two orthogonal comparisons regarding the trend of the trial means. The linear comparison was significant, whereas the quadratic was not.

Drugs × Trials Interaction

We examine now the differences between the linear components of the trends for the three drug groups. By substitution in formula (14.2) we have

$$\begin{aligned}\text{Linear component} &= \frac{(30)^2 + (20)^2 + (20)^2}{(5)(2)} - \frac{(70)^2}{(3)(5)(2)} \\ &= 170.00 - 163.33 \\ &= 6.67\end{aligned}$$

with $k - 1 = 2$ d.f. The mean square is thus $6.67/2 = 3.33$. For the test of significance we have $F = 3.33/.83 = 4.01$ with 2 and 24 d.f. This value of F is significant at the 5 per cent level and indicates that the linear components of the trends for the three drugs differ significantly. It is apparent from Figure 14.1 that the trends for A_1 and A_3 are exactly linear. For each of these two groups, the linear components of the trends will be exactly equal to the corresponding sums of squares between trials. Thus, the quadratic components of the trends for both groups, as shown in Table 14.14, are zero.

For the quadratic component, representing differences in curvature of the trends, we have

$$\begin{aligned}\text{Quadratic component} &= \frac{(0)^2 + (10)^2 + (0)^2}{(5)(6)} - \frac{(10)^2}{(3)(5)(6)} \\ &= 3.33 - 1.11 \\ &= 2.22\end{aligned}$$

with $k - 1 = 2$ d.f. The mean square is thus $2.22/2 = 1.11$. We have $F = 1.11/.83 = 1.34$, a nonsignificant value for 2 and 24 d.f. We conclude that the differences in curvature of the trends are not significant.

We observe that the sum of squares for differences between the linear components is 6.67 and the sum of squares for differences in the quadratic components is 2.22. Each of these sums of squares has 2 d.f., and their sum is equal to 8.89 with 4 d.f. This, of course, is the sum of squares for the drugs $\times$ trials interaction. In our discussion of the analysis of the experiment in the previous section, we stated that the sums of squares given by formulas (14.2) and (14.3), as these equations were used in the analysis, were components of interaction sums of squares. In the present example, we see that this is so. In essence, we have analyzed the drugs $\times$ trials sum of squares, with 4 d.f., into two orthogonal comparisons, each with 2 d.f. One of these corresponds to differences in the linear components of the group trends and the other to differences in the quadratic components.

Note on the Error Sum of Squares

Suppose, in the present experiment, we also multiply the scores of each subject in each group by the linear coefficients. For the subjects tested with Drug A_1, we would have

$$\begin{aligned}
(-1)(2) + (0)(4) + (1)(7) &= 5 \\
(-1)(2) + (0)(6) + (1)(10) &= 8 \\
(-1)(3) + (0)(7) + (1)(10) &= 7 \\
(-1)(7) + (0)(9) + (1)(11) &= 4 \\
(-1)(6) + (0)(9) + (1)(12) &= 6
\end{aligned}$$

Then, by formula (14.2), we have

$$\frac{(5)^2 + (8)^2 + (7)^2 + (4)^2 + (6)^2}{(1)(2)} - \frac{(30)^2}{(5)(1)(2)} = 5.0$$

as the linear component of the S's $\times$ trials sum of squares for A_1 with 4 d.f. Similarly, for the group tested with A_2, we would have

$$\frac{(5)^2 + (6)^2 + (4)^2 + (3)^2 + (2)^2}{(1)(2)} - \frac{(20)^2}{(5)(1)(2)} = 5.0$$

and for the group tested with A_3 we would have

$$\frac{(4)^2 + (6)^2 + (5)^2 + (2)^2 + (3)^2}{(1)(2)} - \frac{(20)^2}{(5)(1)(2)} = 5.0$$

Each of the above sums of squares has 4 d f. and the pooled sum of squares is 15 0 with 12 d.f. This is the linear component of the pooled S's $\times$ trials sum of squares.

To obtain the quadratic component of the pooled S's $\times$ trials sum of squares we would proceed in the same way using the quadratic coefficients. Doing this for each of the three groups separately, we have

$$\frac{(1)^2 + (0)^2 + (-1)^2 + (0)^2 + (0)^2}{(1)(6)} - \frac{(0)^2}{(5)(1)(6)} = .33$$

$$\frac{(3)^2 + (4)^2 + (2)^2 + (1)^2 + (0)^2}{(1)(6)} - \frac{(10)^2}{(5)(1)(6)} = 1.67$$

$$\frac{(2)^2 + (0)^2 + (-1)^2 + (2)^2 + (-3)^2}{(1)(6)} - \frac{(0)^2}{(5)(1)(6)} = 3.00$$

Each of the above sums of squares has 4 d.f. and the pooled sum of squares is 5.00 with 12 d.f.

We note that the sum of the linear component and the quadratic component is 15.00 + 5.00 = 20.00 and this is the pooled S's $\times$ trials sum of squares. In our tests of significance we have assumed that the linear and quadratic components of S's $\times$ trials are estimates of the same common error variance. The calculations above illustrate, however, that it is possible to partition the error sum of squares (S's $\times$ trials) in the same way that we partition the other sums of squares into linear and quadratic components if there should be any serious question concerning the homogeneity of these components.

QUESTIONS AND PROBLEMS

1. Learning scores for five subjects on each of four trials are given below:

Subjects	Trials			
	1	2	3	4
1	2	4	5	7
2	3	5	6	8
3	5	6	8	10
4	7	9	11	14
5	6	7	10	11

(*a*) Analyze the results by the usual analysis of variance methods. (*b*) Determine whether the linear component of the trend of the trial means is significant. (*c*) Determine whether there is significant curvature in the trend.

2. Fifteen subjects were randomly assigned to one of three treatment groups. The measures given below are recall scores for verbal material based upon recall after 1, 2, and 3 days.

	Subjects	Days		
		1	2	3
A_1	1	28	25	22
	2	32	29	24
	3	36	35	27
	4	45	42	40
	5	46	43	40
A_2	1	27	24	20
	2	29	26	22
	3	36	36	30
	4	42	43	39
	5	48	44	40
A_3	1	40	38	33
	2	36	26	20
	3	50	48	44
	4	45	43	35
	5	42	39	30

(*a*) Analyze the results by the usual analysis of variance methods. (*b*) Test the linear component of the over-all trend for significance. Determine whether there is significant curvature in the over-all trend. Note that the sum of the sums of squares for the linear and quadratic components is equal to the sum of squares for days. (*c*) Find the linear and quadratic components of the treatments × days sum of squares. Note that the sum of these two components is equal to the treatments × days interaction sum of squares.

3. Let A_1 and A_2 correspond to two levels of anxiety. Within each level subjects are randomly assigned to two treatments B_1 and B_2. Each subject is then given 4 trials, C_1, C_2, C_3, and C_4. The outcomes of the experiment are as follows:

		C_1	C_2	C_3	C_4
A_1	B_1	7	9	10	11
		8	9	11	12
		8	10	11	12
	B_2	5	6	7	8
		5	6	7	9
		3	4	9	10
A_2	B_1	6	7	8	9
		7	8	10	11
		3	6	9	11
	B_2	2	3	4	5
		3	4	5	7
		2	4	5	7

cance. (*c*) Find the linear and quadratic components of the interaction sum of squares and test each for significance.

5. In the discussion of the analysis of the data of Table 14.3, the pooled S's $\times$ trials sum of squares was obtained by subtraction and was equal to 20.00 with 24 d.f. Calculate the S's $\times$ trials sum of squares for each of the three drugs. Each of these sums of squares will have 8 d.f. The sum of the three sums of squares should also be equal to 20.00.

15

LATIN SQUARE DESIGNS

INTRODUCTION

In our discussion of the randomized blocks design, the blocks were considered as representing a particular group of subjects. The blocks were formed on the basis of prior measurements or prior knowledge of the subjects. In forming blocks, it is anticipated that, in the absence of treatment effects, variation within each block will, in general, be less than would be found if the same number of observations were obtained from subjects selected completely at random. If there is substantial variation between the blocks, the analysis of variance effectively eliminates this source of variation from the estimate of experimental error.

Consider now a situation where the laboratory technique is such that only 5 subjects can be tested each day and we have 5 treatments, A, B, C, D, and E. We have no prior measurements or knowledge about the subjects that we can use as a guide in arranging the subjects in blocks. But assume that the variation between observations made on different days is of some importance. Then, a group of 5 subjects, randomly assigned to each day's testing, could be considered as forming a block. Within each day, the 5 treatments could be assigned at random. Thus we might have the following randomization:

Day 1	A	B	C	E	D
Day 2	C	B	E	A	D
Day 3	A	C	D	B	E
Day 4	B	A	C	D	E
Day 5	A	B	C	E	D

The analysis of variance for this experiment would be the same as that for a randomized blocks design, with the days corresponding to blocks. If the day-to-day variation is of some importance, then the design will remove this source of variation from the estimate of experimental error.

It may occur to the experimenter, however, that another source of variation may be involved which also should be eliminated from the estimate of experimental error. Suppose, for example, that one of the 5 subjects is tested each day at the following hours: 1, 2, 3, 4, and 5 o'clock. The

Table 15.1 Example of a 5 × 5 Latin Square

Days	Hours				
	1	2	3	4	5
Monday	*B*	*E*	*D*	*C*	*A*
Tuesday	*C*	*A*	*B*	*E*	*D*
Wednesday	*D*	*B*	*C*	*A*	*E*
Thursday	*E*	*C*	*A*	*D*	*B*
Friday	*A*	*D*	*E*	*B*	*C*

variation between the hours may be of importance and might be controlled by restricting the randomization in such a way that each treatment occurs not only once on each day but also only once at each hour. With this restricted randomization we may have the arrangement of the treatments shown in Table 15.1.

In Table 15.1, it may be observed that each treatment (letter) occurs once and only once in each row and each column. An arrangement of this kind is called a *Latin square.* To form a Latin square we must have the number of rows equal to the number of columns equal to the number of treatments. We shall consider first the analysis of variance for a Latin square design and then the problem of randomization.

ANALYSIS OF VARIANCE OF THE LATIN SQUARE DESIGN

Assume that, in the experiment described, the observations corresponding to the cell entries are those shown in Table 15.2. The total sum

Table 15.2 Observations Obtained with the Latin Square Design of Table 15.1

Days	Hours					Σ
	1	2	3	4	5	
Monday	8	18	5	8	6	45
Tuesday	1	6	5	18	9	39
Wednesday	5	4	4	8	14	35
Thursday	11	4	14	1	7	37
Friday	9	9	16	3	2	39
Σ	34	41	44	38	38	195

of squares based upon the variation of the cell entries of this table may be calculated in the usual manner. Thus, we have

$$\text{Total} = (8)^2 + (1)^2 + \cdots + (2)^2 - \frac{(195)^2}{25} = 590.0$$

The sum of squares between hours (columns) will be given by

$$\text{Columns} = \frac{(34)^2}{5} + \frac{(41)^2}{5} + \cdots + \frac{(38)^2}{5} - \frac{(195)^2}{25} = 11.2$$

The sum of squares between days (rows) will be given by

$$\text{Rows} = \frac{(45)^2}{5} + \frac{(39)^2}{5} + \cdots + \frac{(39)^2}{5} - \frac{(195)^2}{25} = 11.2$$

The treatment sums are obtained by adding the cell entries for each treatment. Thus we have

$$\begin{aligned} A &= 9 + 6 + 14 + 8 + 6 = 43 \\ B &= 8 + 4 + 5 + 3 + 7 = 27 \\ C &= 1 + 4 + 4 + 8 + 2 = 19 \\ D &= 5 + 9 + 5 + 1 + 9 = 29 \\ E &= 11 + 18 + 16 + 18 + 14 = 77 \end{aligned}$$

Then the sum of squares for treatments will be equal to

$$\text{Treatments} = \frac{(43)^2}{5} + \frac{(27)^2}{5} + \cdots + \frac{(77)^2}{5} - \frac{(195)^2}{25} = 420.8$$

If we subtract the sum of squares for days (rows), hours (columns), and treatments, each with 4 d.f., from the total sum of squares, we obtain a residual sum of squares with 12 d.f. This residual sum of squares is the error sum of squares for the Latin square design. Thus

$$\text{Error} = \text{Total} - \text{rows} - \text{columns} - \text{treatments} \qquad (15.1)$$

For the data of Table 15.2, we have

$$\text{Error} = 590.0 - 11.2 - 11.2 - 420.8 = 146.8$$

The summary of the analysis of variance is shown in Table 15.3. To test the mean square for treatments for significance, we have $F =$

Table 15.3 Analysis of Variance for the Data of Table 15.2

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Treatments	420 8	4	105.20	8.60
Days	11.2	4	2.80	
Hours	11.2	4	2 80	
Error	146 8	12	12.23	
Total	590.0	24		

105.20/12.23 = 8.60 with 4 and 12 d f. This is a significant value and we conclude that the treatment means differ significantly. In the absence of planned comparisons, we may apply Duncan's multiple range test to the treatment means, using as our estimate of s^2 the error mean square of the analysis of variance which has 12 d.f.

GENERAL EQUATION FOR THE LATIN SQUARE

We obtained the error sum of squares by subtraction. It can, however, be calculated directly. For the Latin square, we have r rows, c columns, and t treatments. Let the observation in the rth row, the cth column, with the tth treatment be designated by X_{rct}, with the understanding that when used as subscripts, r, c, and t correspond to variables Then the deviation of any observation from the over-all mean can be expressed as

$$\begin{aligned} X_{rct} - \bar{X}... = {} & (\bar{X}_{r}.. - \bar{X}...) \\ & + (\bar{X}_{.c}. - \bar{X}...) \\ & + (\bar{X}_{..t} - \bar{X}...) \\ & + (X_{rct} - \bar{X}_{r}.. - \bar{X}_{.c}. - \bar{X}_{..t} + 2\bar{X}...) \end{aligned}$$

The successive terms on the right correspond to the deviation of the row mean from the over-all mean, the deviation of the column mean from the over-all mean, the deviation of the treatment mean from the over-all mean, and the last term is a residual. If we squared the above expression and summed over all rc observations we would find that all sums of products between the terms on the right sum to zero. We have $r = c = t = k$, with a total of rc observations. Then

$$\begin{aligned} \sum_{1}^{rc} (X_{rct} - \bar{X}...)^2 = {} & k \sum_{1}^{r} (\bar{X}_{r}.. - \bar{X}...)^2 \\ & + k \sum_{1}^{c} (\bar{X}_{.c}. - \bar{X}...)^2 \\ & + k \sum_{1}^{t} (\bar{X}_{..t} - \bar{X}...)^2 \\ & + \sum_{1}^{rc} (X_{rct} - \bar{X}_{r}.. - \bar{X}_{.c}. - \bar{X}_{..t} + 2\bar{X}...)^2 \end{aligned}$$

The term on the left of the above expression gives the total sum of squares. The successive terms on the right give the sums of squares for rows, columns, treatments, and error.

RANDOMIZATION PROCEDURES

We turn now to the problem of randomization.[1] Table 15.4 lists selected Latin square arrangements for squares of order 3 to 7.[2] Assume that we have 4 treatments. Select at random one of the 4 × 4 squares. We can do this by drawing at random a number from 1 to 4. We then randomize the rows and columns of the square and assign the treatments at random to the

Table 15.4 Examples of Selected Latin Squares

3 × 3	4 × 4			
	(a)	(b)	(c)	(d)
A B C	A B C D	A B C D	A B C D	A B C D
B C A	B A D C	B C D A	B D A C	B A D C
C A B	C D B A	C D A B	C A D B	C D A B
	D C A B	D A B C	D C B A	D C B A

5 × 5	6 × 6	7 × 7
A B C D E	A B C D E F	A B C D E F G
B A E C D	B C F A D E	B C D E F G A
C D A E B	C F B E A D	C D E F G A B
D E B A C	D E A B F C	D E F G A B C
E C D B A	E A D F C B	E F G A B C D
	F D E C B A	F G A B C D E
		G A B C D E F

letters. Suppose we have randomly selected square (a) from the set of 4 squares. Using a table of random numbers, we write down three random permutations of the numbers 1, 2, 3, and 4. Suppose our point of entry into Table I is the first block, row 01, and column 10. Reading down, the first permutation we obtain is: 1, 3, 4, 2. Continuing, we obtain 4, 1, 2, 3, and 2, 4, 3, 1. We rearrange the rows of square (a) in accordance with the first permutation. This gives us

1	*A*	*B*	*C*	*D*
3	*C*	*D*	*B*	*A*
4	*D*	*C*	*A*	*B*
2	*B*	*A*	*D*	*C*

[1] For a more exact and complete discussion of procedures for selecting a Latin square at random, see Fisher and Yates (1948).

[2] Squares larger than 7 × 7 can be constructed in the same manner as the 7 × 7 square given in Table 15.4 To construct an 8 × 8 square, write the first row $A \cdots H$, the second row $B \cdots HA$, the third row $C \cdots HAB$, the fourth row $D \cdots HABC$, and so on.

We now rearrange the columns of the above square in accordance with the second permutation. This gives us

4	1	2	3
D	*A*	*B*	*C*
A	*C*	*D*	*B*
B	*D*	*C*	*A*
C	*B*	*A*	*D*

Then, if the treatments have been numbered 1, 2, 3, and 4, we rearrange them in accordance with the third permutation, and this permutation is used to assign the letters of the Latin square to the treatments. Thus we would have

Treatment Number:	1	2	3	4
Permutation:	2	4	3	1
Letter:	*A*	*B*	*C*	*D*

Thus Treatment 2 will be assigned A; Treatment 4, B; Treatment 3, C; and Treatment 1, D in the Latin square.

REPLICATION WITH INDEPENDENT SQUARES

If we use a Latin square design with t treatments, then the error mean square will have $(t - 1)(t - 2)$ degrees of freedom. For the 5×5 square, for example, we have 12 d.f. for the error mean square. The error sum of squares for the 6×6 Latin square will have 20 d.f. and the 7×7 square results in an error sum of squares with 30 d.f. If the estimate of experimental error based upon less than 30 d.f. is not very reliable, then it is clear that the smaller Latin squares will not be useful. To obtain an estimate with a larger number of degrees of freedom, we can replicate the complete experiment with additional Latin squares.[3]

The Bliss and Rose Experiment

In the applications to which the Latin square design has been typically put in psychology, physiology, and drug research, each row or block of the square has consisted of a single subject with the columns corresponding to

[3] This is not the only reason for having additional replications Other things being equal, if the true treatment means differ, then, regardless of the nature of the experimental design, increasing the number of replications for each of the treatments will also serve to increase the treatment mean square. Thus, if there are true treatment differences, they are more apt to be detected, declared significant, as the number of replications is increased. See, for example, the discussion of expectations of mean squares in Chapter 17.

successive periods or times. For example, in an experiment by Bliss and Rose (1940) a 4 × 4 Latin square was used. The treatments consisted of a dosage of an extract of parathyroid and 4 different preparations were used. The experiment was one in biological assay in which two dosages, U_1 and U_2, of an unknown preparation of parathyroid extract were administered along with two dosages, S_1 and S_2, of a standard preparation. The dosages were given to each of 4 dogs at 4 different times. Thus, in this instance, each row of the square corresponds to a single subject with the columns corresponding to successive periods or tests. It is this kind of application of the Latin square that occurs most frequently in physiological and psychological research.

Table 15.5 Milligrams Per Cent Calcium Secretion of Dogs Given Four Different Preparations of Parathyroid Extract

Latin Squares	Dogs	Days				
		3/15	3/25	4/5	4/15	Σ
S_1 S_2 U_2 U_1	1	13.8	17 0	16.0	16.0	62.8
U_2 U_1 S_1 S_2	2	15.8	14 3	14.8	15.4	60 3
S_2 S_1 U_1 U_2	3	15.0	14.5	14.0	15.0	58.5
U_1 U_2 S_2 S_1	4	14.7	15.4	14.8	14.0	58.9
	Σ	59.3	61.2	59.6	60.4	240.5
U_2 U_1 S_1 S_2	5	17.0	16.5	15.0	15.4	63.9
U_1 U_2 S_2 S_1	6	15.1	15.0	15.8	13.4	59.3
S_2 S_1 U_1 U_2	7	15.0	14.0	14.6	15.6	59.2
S_1 S_2 U_2 U_1	8	12.0	13.8	14.0	13.8	53.6
	Σ	59.1	59.3	59.4	58.2	236.0
S_2 U_2 S_1 U_1	9	14.6	15.4	14.0	14.8	58.8
U_1 S_1 U_2 S_2	10	13.6	15.3	17.2	15.3	61.4
U_2 S_2 U_1 S_1	11	14.4	13.8	14.4	15.0	57.6
S_1 U_1 S_2 U_2	12	15.8	15.0	15.2	15.8	61.8
	Σ	58.4	59.5	60.8	60.9	239.6
U_1 U_2 S_1 S_2	13	14.0	13.8	14.0	14.0	55.8
U_2 U_1 S_2 S_1	14	16.2	14.0	13.0	13.0	56.2
S_2 S_1 U_2 U_1	15	13.0	14.0	14.0	13.0	54.0
S_1 S_2 U_1 U_2	16	13.2	16.0	14.9	16.4	60.5
	Σ	56.4	57.8	55.9	56.4	226.5
S_1 U_1 S_2 U_2	17	14.2	14.1	15.0	14.4	57.7
U_1 S_1 U_2 S_2	18	13.0	13.4	13.8	14.0	54.2
U_2 S_2 U_1 S_1	19	15.8	16.0	15.0	15.4	62 2
S_2 U_2 S_1 U_1	20	15.2	16.2	15.0	15.3	61.7
	Σ	58.2	59.7	58.8	59.1	235.8

Since the estimate of experimental error for the 4×4 square has only 6 d.f., the complete experiment was replicated by using 5 Latin squares. Each of these squares was independently randomized in the manner previously described. In the experiment, 20 dogs were divided at random into 5 sets of 4 each. For each set of dogs an independently randomized square was used. On the first day, all 20 dogs were given the treatment prescribed by the separate Latin square entries corresponding to the first column of all 5 Latin squares. At the time of the second test all animals were given the treatment prescribed by the separate Latin square entries corresponding to the second column, and so on. The observations recorded in Table 15.5 consist of the milligrams per cent calcium secretion of the dogs for each treatment.

Analysis of Variance for a Single Square

If we consider only the first Latin square of Table 15.5, we obtain the following sums of squares:

$$\text{Total} = (13.8)^2 + (15.8)^2 + \cdots + (14.0)^2 - \frac{(240.5)^2}{16} = 11.294$$

$$\text{Days} = \frac{(59.3)^2}{4} + \frac{(61.2)^2}{4} + \cdots + \frac{(60.4)^2}{4} - \frac{(240\ 5)^2}{16} = .546$$

$$\text{Dogs} = \frac{(62.8)^2}{4} + \frac{(60.3)^2}{4} + \cdots + \frac{(58.9)^2}{4} - \frac{(240.5)^2}{16} = 2.832$$

Table 15.6 Analysis of Variance for the First Latin Square of Table 15.5

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Treatments	4.756	3	1.585	3.01
Dogs	2.832	3	.944	
Days	.546	3	.182	
Error	3.160	6	.527	
Total	11.294	15		

Summing the entries for each preparation we obtain the following sums:

$$S_1 = 57.1, \quad S_2 = 62.2, \quad U_1 = 59.0, \quad \text{and } U_2 = 62.2$$

Then, the sum of squares for drugs will be

$$\text{Drugs} = \frac{(57.1)^2}{4} + \frac{(62.2)^2}{4} + \cdots + \frac{(62.2)^2}{4} - \frac{(240.5)^2}{16} = 4.756$$

and for the error sum of squares we have

$$\text{Error} = 11.294 - .546 - 2.832 - 4.756 = 3.160$$

The analysis of variance for the first square is shown in Table 15.6. Testing the treatment mean square for significance, we have $F = 1.585/.527 = 3.01$ with 3 and 6 d.f. From the table of F we find that $F = 3.01$ is a nonsignificant value.

Analysis of Variance for Replications with Independent Squares

Now a similar analysis can be performed with each of the separate Latin squares. If we then add together the corresponding sums of squares and degrees of freedom, we obtain the analysis shown in Table 15.7. The next to last entry in the table is a sum of squares between the 5 Latin

Table 15.7 Pooled Sums of Squares and Degrees of Freedom for the Five Latin Squares of Table 15.5

Source of Variation	Sum of Squares	d f.
Treatments	22.57	15
Dogs	35.50	15
Days	2.62	15
Error	18.21	30
Latin Squares	7.69	4
Total	86.59	79

squares with 4 d.f., and can be obtained by calculating

$$\text{Squares} = \frac{(240.5)^2}{16} + \frac{(236.0)^2}{16} + \cdots + \frac{(235.8)^2}{16} - \frac{(1{,}178.4)^2}{90} = 7.69$$

The sum of squares for days in Table 15.7 is composed of two component parts. For example, consider Table 15.8. If we calculated the sum of squares between the cells of the table, subtracted the row and column sums of squares from the between-cells sum of squares, we would have the rows $\times$ columns interaction. In the present instance, the rows $\times$ columns sum of squares would be the squares $\times$ days sum of squares and would have 12 d.f. The column sum of squares is the over-all day sum of squares

Table 15.8 Two-Way Table for Squares and Days

	Days				Σ
	1	2	3	4	
Square 1	59.3	61.2	59 6	60.4	240.5
Square 2	59.1	59.3	59.4	58.2	236.0
Square 3	58.4	59.5	60.8	60.9	239.6
Square 4	56.4	57.8	55.9	56.4	226.5
Square 5	58.2	59.7	58.8	59.1	235 8
Σ	291 4	297.5	294.5	295.0	1,178 4

with 3 d.f. Making the necessary calculations we find that the squares × days sum of squares is equal to 1.68 and the day sum of squares is equal to .94. The sum of these two sums of squares is 2.62 and this is the value shown in Table 15.7 for days with 15 d f.

Just as we did in the case of the two-way table for squares and days, we can set up a two-way table for squares and drugs, as shown in Table 15.9. The cell entries of Table 15.9 would represent the sums for each drug

Table 15.9 Two-Way Table for Squares and Drugs

	Drugs			
Square 1	S_1	S_2	U_1	U_2
Square 2	S_1	S_2	U_1	U_2
Square 3	S_1	S_2	U_1	U_2
Square 4	S_1	S_2	U_1	U_2
Square 5	S_1	S_2	U_1	U_2

for each square. The over-all column sum of squares would be the sum of squares for drugs with 3 d.f., and the rows × columns interaction sum of squares would be the squares × drugs interaction with 12 d.f. In Table 15.7 the sum of squares for drugs (treatments) consists of both of these components. By making the necessary calculations, we find that the squares × drugs sum of squares is equal to 7.42 and the sum of squares for drugs is equal to 15.15.

Summary of the Analysis

In Table 15.10 we show the summary of the analysis. Testing the treatment mean square for significance we have $F = 5.050/.607 = 8.32$ with 3 and 30 d.f. From the table of F we find that $F = 8.32$ is a significant value. Additional tests on the treatment means may be made using the procedures described previously for making multiple comparisons.

It is possible to consider each of the separate Latin squares as a repetition or replication of the complete experiment. In this sense, each Latin square is a block and the squares × drugs sum of squares is similar to a

Table 15.10 Analysis of Variance for the Data of Table 15.5

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Squares	7.69	4	1.922	
Between dogs in same square	35.50	15	2.367	
Treatments	15.15	3	5.050	8.32
Days	.94	3	.313	
Squares × days	1.68	12	.140	
Squares × treatments	7.42	12	.618	
Error	18.21	30	.607	
Total	86.59	79		

blocks × treatments sum of squares in a randomized blocks design. As in the case of a randomized blocks design, under the assumption that the squares × drugs interaction is negligible, this mean square may also be considered as an estimate of experimental error. It is apparent in Table 15.10 that the squares × drugs mean square (.618) and the error mean square (.607) are comparable. Therefore, as our estimate of experimental error, we might take the sum of squares $7.42 + 18.21 = 25.63$ with $12 + 30 = 42$ d.f. Then the error mean square for the analysis would be $25.63/42 = .610$

Similarly, we may also pool the squares × columns (days) sum of squares as well as the squares × treatments sum of squares with the error sum of squares. In doing so, the assumption is made that the column effects are approximately the same from square to square. This is another way of stating that the squares × columns interaction is assumed to be negligible so that the squares × columns mean square is also an estimate of experimental error. Pooling the squares × columns sum of squares with the squares × treatments sum of squares and the error sum of squares, we have $1.68 + 7.42 + 18.21 = 27.31$ with $12 + 12 + 30 = 54$ d.f. Then the error mean square would be $27.31/54 = .506$.

Before combining the results from separate Latin squares in a single analysis, it is advisable first to analyze the data of the separate Latin squares. If the error mean squares of the separate Latin squares vary considerably, then this may be a sign that the experimental technique is not under control. In the present example, Bartlett's test for homogeneity of variance shows that the 5 estimates of experimental error obtained from the separate squares do not differ significantly.

With a reliable experimental technique, one might anticipate that the differences between the treatment means in each square would be fairly comparable from square to square. If this is the case, then the squares × treatments mean square should not be significantly greater than the error mean square and we may combine the two sums of squares and their corresponding degrees of freedom to obtain the estimate of experimental error.[4] Similarly, unless we have some a priori reason to believe that the differences between the column means in each square will not be fairly comparable from square to square, we would ordinarily combine the squares × columns sum of squares with the squares × treatments sum of squares and the error sum of squares to obtain a single estimate of experimental error.

[4] If the squares have been randomly selected from a larger population of possible squares, and if the squares × treatments mean square should be significantly larger than the error mean square, then the appropriate error term for testing the treatment mean square for significance is the squares × treatments mean square. See, for example, the article by Wilk and Kempthorne (1957).

REPLICATION OF THE SAME SQUARE

Nature of the Experiment

Let us now consider the data of Table 15.11. The experiment was concerned with the ability of subjects to locate targets when they appeared on circular screens of varying size. Screen sizes of 3, 4, 5, 6, and 7 inches

Table 15.11 Number of Correct Judgments for Each Subject for Each Screen Size

Latin Square	Subjects	Periods					Σ
		1	2	3	4	5	
	1	11	15	10	13	10	59
	2	7	18	12	17	16	70
3 6 4 7 5	3	12	17	9	19	18	75
	4	13	16	17	15	15	76
	5	14	17	18	18	15	82
	Σ	57	83	66	82	74	362
	6	8	11	11	15	14	59
	7	13	13	13	5	10	54
4 7 5 3 6	8	19	16	18	14	21	88
	9	8	17	12	14	17	68
	10	9	9	16	10	11	55
	Σ	57	66	70	58	73	324
	11	10	8	13	11	17	59
	12	16	16	18	14	19	83
5 3 6 4 7	13	15	15	16	14	14	74
	14	12	14	15	13	16	70
	15	12	10	14	12	16	64
	Σ	65	63	76	64	82	350
	16	13	12	16	14	9	64
	17	16	10	19	16	11	72
6 4 7 5 3	18	13	15	13	13	9	63
	19	11	10	13	11	7	52
	20	6	16	13	17	10	62
	Σ	59	63	74	71	46	313
	21	11	8	9	12	15	55
	22	20	18	10	18	17	83
7 5 3 6 4	23	16	14	10	12	13	65
	24	14	15	16	16	17	78
	25	11	12	10	15	15	63
	Σ	72	67	55	73	77	344

were used in the experiment. Radii were marked on the screens at 20 degree intervals. Each screen was also marked by a series of expanding circles representing intervals of 10 miles from the center of the target. The screens were exposed to the subject by means of an automatic timer at the rate of one screen every 15 seconds. The screens had been photographed on a film strip and were projected to a round glass plate. There were 36 projections for each of the 5 screen sizes and the subjects were asked to locate the position of a target which appeared on the screen in terms of both degrees and miles. The data given in Table 15.11 are the number of correct judgments for degrees only.

With 5 subjects a single Latin square may be formed with the rows corresponding to subjects, the columns to successive test periods, and the cell entries to the screen sizes. In accordance with a Latin square design, each screen size would appear once and only once in each column and each row. Instead of replicating the experiment with additional and independently randomized Latin squares, the experimenter chose to replicate the same Latin square 5 times with 5 subjects assigned at random to each row of the square. This particular design makes possible the isolation of a sum of squares corresponding to the particular sequences or orders of presentation of the screen sizes, although we must keep in mind that only 5 of the possible 120 different orders were investigated.

Sums of Squares

For the data of Table 15.11, we find the following sums of squares:

$$\text{Total} = (11)^2 + (7)^2 + \cdots + (15)^2 - \frac{(1{,}693)^2}{125} = 1{,}327.01$$

$$\text{Rows} = \frac{(59)^2}{5} + \frac{(70)^2}{5} + \cdots + \frac{(63)^2}{5} - \frac{(1{,}693)^2}{125} = 495.41$$

Then from Table 15.11 we form Table 15.12 where each cell entry is the sum of 5 observations. For this table, we find the sum of squares for rows, corresponding to the 5 orders or sequences of screen sizes. Thus

$$\text{Orders} = \frac{(362)^2}{25} + \frac{(324)^2}{25} + \cdots + \frac{(344)^2}{25} - \frac{(1{,}693)^2}{125} = 63.01$$

This sum of squares is part of the row sum of squares, 495.41. Subtracting the order sum of squares from the sum of squares between rows, we have a residual, $495.41 - 63.01 = 432.40$ with $24 - 4 = 20$ d.f. As should be clear from previous analyses, the residual is the pooled sum of squares between subjects tested with the same order. For example, the sum of squares between subjects, with order 3, 6, 4, 7, 5, would have $5 - 1 = 4$ d.f. Calculating this sum of squares for each order, we would have 5 such sums

Table 15.12 Two-Way Table for Orders and Periods

Order	Periods					Σ
	1	2	3	4	5	
3–6–4–7–5	57	83	66	82	74	362
4–7–5–3–6	57	66	70	58	73	324
5–3–6–4–7	65	63	76	64	82	350
6–4–7–5–3	59	63	74	71	46	313
7–5–3–6–4	72	67	55	73	77	344
Σ	310	342	341	348	352	1,693

of squares with $(5)(5-1) = 20$ d.f. This is the sum of squares on which error mean square (a) of Table 15.13 is based.

From Table 15.12, we also calculate the sum of squares for periods and this is equal to

$$\text{Periods} = \frac{(310)^2}{25} + \frac{(342)^2}{25} + \cdots + \frac{(352)^2}{25} - \frac{(1{,}693)^2}{125} = 44.13$$

From this table we also find the sums for each screen size. For example, the sum for the 3-inch screen is equal to $57 + 58 + 63 + 46 + 55 = 279$. The sums for the 4-, 5-, 6-, and 7-inch screens are 327, 347. 364, and 376, respectively. Then the sum of squares for screens will be

$$\text{Screens} = \frac{(279)^2}{25} + \frac{(327)^2}{25} + \cdots + \frac{(376)^2}{25} - \frac{(1{,}693)^2}{125} = 232.05$$

If we now find the sum of squares between the cells of Table 15.12, we can then subtract the sums of squares for orders, screens, and periods to obtain the Latin square error sum of squares. Thus

$$\text{Between cells} = \frac{(57)^2}{5} + \frac{(57)^2}{5} + \cdots + \frac{(77)^2}{5} - \frac{(1{,}693)^2}{125} = 414.21$$

and by subtraction, we have

$$\text{Error} = 414.21 - 63.01 - 44.13 - 232.05 = 75.02$$

For each of the 5 orders we have 5 subjects tested at 5 different periods. It is thus possible to find a sum of squares for each order which is the S's $\times$ periods interaction for that order. To actually calculate this sum of squares for the first order we would find the sum of squares between the 25 observations for the first order. From this sum of squares we would then subtract the sum of squares between subjects (rows) and the sum of squares between periods (columns). The residual is the S's $\times$ periods interaction

for the first order with 16 d.f. Repeating these calculations for each order, we would have 5 different S's $\times$ periods interactions. The sum of these sums of squares is equal to 480.40 with 80 d.f. If we are interested only in the *pooled* S's $\times$ periods interaction with 80 d.f., it should be clear that this sum of squares can be obtained by subtraction.

Summary of the Analysis

Under usual circumstances, we expect the Latin square error to be an estimate of the same error as the S's $\times$ periods mean square. Therefore, in Table 15.13 we have pooled the Latin square error sum of squares and degrees of freedom with those for S's $\times$ periods, to obtain the error mean square designated (b) with 92 d.f.[5] This is the appropriate error term for

Table 15.13 Analysis of Variance for the Data of Table 15.11

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Orders	63.01	4	15.75	
Error (a)	432.40	20	21.62	
Screens	232.05	4	58.01	9.60
Periods	44.13	4	11.03	1.83
Error (b)[1]	555.42	92	6.04	
Total	1,327.01	124		

[1] The sum of squares for error (b) is the sum of the Latin square error sum of squares (75.02) and the pooled S's $\times$ periods sum of squares (480.40) with 12 and 80 d.f., respectively.

testing the significance of screen sizes and periods. The error mean square designated (a) is the appropriate error term for testing the order mean square. We note in this analysis, as in those discussed previously, error (b) is smaller than error (a) and this will usually be the case.

Trend Analysis for Screen Size

The only significant effect, as shown in Table 15.13, is the screen size. Dividing each of the screen sums by 25, we have the following means for each screen size:

Screen size:	3	4	5	6	7
Mean:	11.16	13.08	13.88	14.56	15.04

The means are plotted against the screen size in Figure 15.1. To determine whether the linear component of the trend is significant, we use the or-

[5] For the Latin square error mean square we have 75.02/12 = 6.25 and for the S's $\times$ periods mean square we have 480 40/80 = 6.00. For the test of homogeneity of the two variances, we have F = 6.25/6.00 = 1.04, with 12 and 80 d.f., a nonsignificant value.

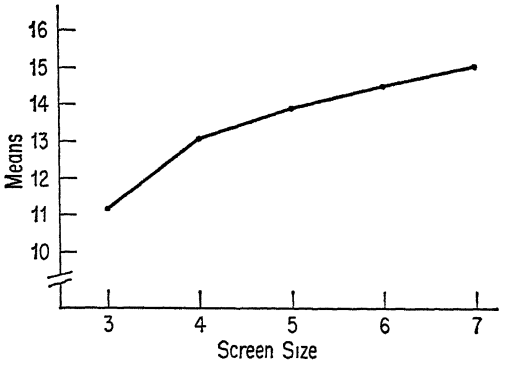


Figure 15.1 Means for five screen sizes.

thogonal coefficients for $k = 5$, obtained from Table XI. Then, multiplying the screen sums by the coefficients, we have

$$D_1 = (-2)(279) + (-1)(327) + (0)(347) + (1)(364) + (2)(376) = 231$$

The sum of the squared orthogonal coefficients is 10, and we have $n = 25$ observations for each screen size. Then the sum of squares for the linear component, as given by formula (10.8), is

$$\frac{(231)^2}{(25)(10)} = 213.44$$

with 1 d.f. For the test of significance, we have $F = 213.44/6.04 = 35.34$ with 1 and 92 d.f. This is a significant value and we conclude that the linear component of the trend is significant.

It would seem from Figure 15.1 that there is a slight curvature in the trend of the means. To test for significance of curvature, we use the orthogonal coefficients for the quadratic component. Then multiplying the sums for the screen sizes by these coefficients, we have

$$D_2 = (2)(279) + (-1)(327) + (-2)(347) + (-1)(364) + (2)(376) = -75$$

The sum of the squared orthogonal coefficients is 14 and we have 25 observations for each sum. Then, by formula (10.8), we have

$$\frac{(-75)^2}{(25)(14)} = 16.07$$

with 1 d.f. For the test of significance of curvature, we have $F = 16.07/6.04 = 2.66$ with 1 and 92 d.f., a nonsignificant value. We conclude, therefore, that the trend of the means is essentially linear and without significant curvature.

LATIN SQUARES AND FRACTIONAL REPLICATION

Suppose we have a $2 \times 2 \times 2$ factorial design and we decide upon ½ fractional replication. As indicated previously, if only ½ of the treatment combinations are to be replicated, then we should choose either the set with plus signs or the set with minus signs in the $A \times B \times C$ comparison. If we take the set with plus signs, then we would replicate each of the following treatment combinations:

$$A_1B_1C_1 \qquad A_1B_2C_2 \qquad A_2B_1C_2 \qquad A_2B_2C_1$$

We may rearrange these treatment combinations in the form of Table 15.14, where each cell entry corresponds to one of the treatment combinations.

Table 15.14 A 2×2 Latin Square

	B_1	B_2
A_1	C_1	C_2
A_2	C_2	C_1

Examination of Table 15.14 shows that the arrangement is that of a 2×2 Latin square for the levels of C, with rows corresponding to the levels of A, and the columns to the levels of B. This 2×2 Latin square, in essence, corresponds to a ½ fractional replication of the $2 \times 2 \times 2$ factorial.

Similarly, consider Table 15.15, where we have a Latin square for 5

Table 15.15 A 5×5 Latin Square

	B_1	B_2	B_3	B_4	B_5
A_1	C_1	C_2	C_3	C_4	C_5
A_2	C_2	C_1	C_5	C_3	C_4
A_3	C_3	C_4	C_1	C_5	C_2
A_4	C_4	C_5	C_2	C_1	C_3
A_5	C_5	C_3	C_4	C_2	C_1

levels of C, with the rows corresponding to the 5 levels of A, and the columns corresponding to the 5 levels of B. For a complete replication of the $5 \times 5 \times 5$ factorial we would have to have 125 observations whereas we have only 25 in Table 15.15. This 5×5 Latin square thus corresponds to a ⅕ fractional replication of the $5 \times 5 \times 5$ factorial.

In typical applications of the Latin square design in the behavioral sciences, the rows or levels of A correspond to subjects and the columns or levels of B correspond to periods, trials, or times of testing. The cell entries or levels of C correspond to treatments. It can be shown that if the $A \times C$ (rows $\times$ treatments) or $B \times C$ (columns $\times$ treatments) inter-

actions of the Latin square are not negligible, they will increase the error mean square. Thus, if the treatment mean square is not significant when compared with the error mean square, this may be because of an inflated error mean square resulting from the presence of interactions of either rows or columns or both with the treatments.[6]

THE 2 × 2 LATIN SQUARE

It should be obvious in the case of a single 2 × 2 Latin square that we cannot obtain any estimate of experimental error. In the absence of interactions, the 2 × 2 square provides estimates of the row, column, and treatment effects, each with 1 d.f. To obtain an estimate of experimental error, we must have additional replications of the 2 × 2 square.

Table 15.16 Five Replications of the 2 × 2 Latin Square

Squares

Subjects	Periods 1	Periods 2
1	*A*	*B*
2	*B*	*A*
3	*A*	*B*
4	*B*	*A*
5	*B*	*A*
6	*A*	*B*
7	*A*	*B*
8	*B*	*A*
9	*A*	*B*
10	*B*	*A*

Observations

Subjects	Periods 1	Periods 2	Σ
1	9	6	15
2	5	8	13
3	10	9	19
4	7	7	14
5	4	7	11
6	8	9	17
7	8	4	12
8	5	11	16
9	5	1	6
10	6	4	10
Σ	67	66	133

	Periods 1	Periods 2	Σ
Square 1	14	14	28
Square 2	17	16	33
Square 3	12	16	28
Square 4	13	15	28
Square 5	11	5	16
Σ	67	66	133

	Treatments *A*	Treatments *B*	Σ
Square 1	17	11	28
Square 2	17	16	33
Square 3	15	13	28
Square 4	19	9	28
Square 5	9	7	16
Σ	77	56	133

[6] The statements made above concerning the rows × treatments and columns × treatments interactions assume that we have a "fixed effects" model in which the rows, columns, and treatments are all regarded as *not* being randomly selected from larger populations. The fixed effects model and other models in the analysis of variance are discussed in Chapter 17.

In Table 15 16, we have 5 replications of the 2 × 2 square. It should be clear in the case of the 2 × 2 square we have only two possible orders, *AB* and *BA*, and it does not make any difference whether we replicate the same square or a series of independently randomized squares. Regardless of which procedure we use, each order will occur an equal number of times.

For the analysis of variance of Table 15.16, we obtain the following sums of squares:

$$\text{Total} = (9)^2 + (5)^2 + \cdots + (4)^2 - \frac{(133)^2}{20} = 114.55$$

$$\text{Rows} = \frac{(15)^2}{2} + \frac{(13)^2}{2} + \cdots + \frac{(10)^2}{2} - \frac{(133)^2}{20} = 64.05$$

$$\text{Columns} = \frac{(67)^2}{10} + \frac{(66)^2}{10} - \frac{(133)^2}{20} = .05$$

$$\text{Treatments} = \frac{(77)^2}{10} + \frac{(56)^2}{10} - \frac{(133)^2}{20} = 22.05$$

As we pointed out previously, we have no Latin square error for the 2 × 2 square. To obtain an estimate of experimental error, we can calculate the squares × periods interaction and the squares × treatments interaction from the two tables shown at the bottom of Table 15.16. Each of these interaction sums of squares will have 4 d.f. If we were to make the necessary calculations, we would find that the sum of these two interaction sums of squares is equal to 28.40 with 8 d.f. We have called this pooled sum of squares the error sum of squares in Table 15.17 which summarizes the analysis.[7] For a test of significance of the treatment mean square, we have

[7] The analysis of variance given in Table 15 17 is for what has been called a *change-over* design by Federer (1955) and a *cross-over* design by Cochran and Cox (1957). If the analysis is done strictly in accordance with the principles of replication of independently randomized Latin squares, the error mean square for testing treatments would be the squares × treatments mean square with 4 d f In the present example, since the squares × periods and the squares × treatments sums of squares are equal, the squares × treatments mean square is also 3.55, but has only 4 d.f. If we should have prior reason to believe that the squares × periods sum of squares will be substantially larger than the squares × treatments sum of squares, the smaller error mean square obtained with the Latin square analysis may offset the loss in degrees of freedom. Without prior reasons to believe that this is the case, the analysis of variance shown is to be preferred, since the error mean square will be based upon twice the degrees of freedom available for the squares × treatments mean square.

A cross-over design can be used with any number of treatments, provided the number of replications is a multiple of the number of treatments For example, with 3 treatments the number of replications must be 3, 6, or 9, and so on. Furthermore, each treatment must occur equally often in each period. This can be accomplished by arranging the treatments in a Latin square. Thus if 3 treatments are arranged in a 3 × 3

$F = 22.05/3.55 = 6.21$ with 1 and 8 d.f. With $\alpha = .05$, this is a significant value of F.

Table 15.17 Analysis of Variance for 5 Replications of the 2 × 2 Latin Square

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Treatments	22.05	1	22.05	6.21
Rows	64.05	9	7.12	
Columns	.05	1	.05	
Error	28.40	8	3.55	
Total	114.55	19		

It is not necessary, except as a check upon our arithmetic, to calculate the squares × periods and the squares × treatments interactions. We can also obtain the error sum of squares of Table 15.17 by subtraction. Thus,

$$\text{Error} = 114.55 - 64.05 - .05 - 22.05 = 28.40$$

It is also possible to rearrange the observations of Table 15.16 in such a way that we obtain Table 15.18. If we calculate the sum of squares for the S's × periods interaction for the first order, AB, and also for the second order, BA, each of these sums of squares would have 4 d.f. The sum of

Table 15.18 The Observations of Table 15.16 Rearranged According to the Order, *AB* or *BA*, of the Treatments

Order	Subjects	Period		Σ
		1	2	
	1	9	6	15
	3	10	9	19
AB	6	8	9	17
	7	8	4	12
	9	5	1	6
	Σ	40	29	69
	2	5	8	13
	4	7	7	14
BA	5	4	7	11
	8	5	11	16
	10	6	4	10
	Σ	27	37	64

Latin square and if we have additional independently randomized replications of the 3 × 3 square, a cross-over analysis is possible. For a further discussion of the difference between the cross-over analysis and the Latin square analysis, see Federer (1955), Cochran and Cox (1957), or Kempthorne (1952).

these two sums of squares will be identical with the error sum of squares of Table 15.17.

The row sum of squares, 64.05, of Table 15.17, with 9 d.f. can be analyzed into a sum of squares for orders with 1 d f., and a pooled sum of squares between subjects tested with the same order with 8 d.f. We can obtain these sums of squares from the data as arranged in Table 15.18. Thus

$$\text{Orders} = \frac{(69)^2}{10} + \frac{(64)^2}{10} - \frac{(133)^2}{20} = 1.25$$

and the pooled sum of squares between subjects tested with the same order will be equal to $64.05 - 1.25 = 62.80$.

By direct calculation, we obtain as the sum of squares between subjects with order AB

$$\frac{(15)^2}{2} + \frac{(19)^2}{2} + \cdots + \frac{(6)^2}{2} - \frac{(69)^2}{10} = 51.4$$

and between subjects with order BA

$$\frac{(13)^2}{2} + \frac{(14)^2}{2} + \cdots + \frac{(10)^2}{2} - \frac{(64)^2}{10} = 11.4$$

and the sum of these two sums of squares is 62.8, the same value we obtained by subtraction. The pooled sum of squares between subjects tested with the same order with 8 d.f. provides the error mean square for testing the order mean square for significance.

CARRY-OVER EFFECTS

A basic condition in the analysis of the Latin square design is that the observations for the different treatments are independent in the sense that the value of an observation for one treatment at a given period is not dependent upon the effects of treatments applied during earlier periods. When this condition is not met, we refer to the treatments as having *carry-over* effects.[8] One way in which the experimenter may hope to eliminate the possibility of carry-over effects is to increase the time interval between the various periods in which the treatments are applied. For example, in the Bliss and Rose experiment, it may be observed that a 10-day interval separated the periods in which the drugs were administered. Obviously, one might expect carry-over effects in drug research if the various drugs being investigated are administered in close succession before the effects of the previous drugs have had an opportunity to wear off. In

[8] As we pointed out earlier, in connection with randomized blocks designs, there is always danger of carry-over effects when the same subjects are given a series of different treatments.

psychological research where each subject is given a series of different treatments the presence of carry-over effects seems most likely. For example, early treatments may produce fatigue effects which carry over to the later treatments. If performance on the same task is measured under different treatments, then practice and learning effects occurring during the early treatments may be expected to carry over to later treatments. Furthermore, if an early treatment is such as to produce a feeling of anxiety, failure, or fear in a subject, then it seems likely that such feelings would influence the subject's motivation and behavior during subsequent treatments.

Balanced Designs

Williams (1949) has suggested a variation of the Latin square arrangement in which each treatment follows every other treatment the same number of times. Latin squares of this kind are often called *balanced* squares. If the number of treatments is even, then for the first row of the balanced square we take

$$1, 2, n, 3, n-1, 4, n-2, 5, n-3, \cdots$$

in which the sequence $1, n, n-1, n-2, n-3, \cdots$ alternates with the sequence $2, 3, 4, 5, \cdots$. For example, with $n = 4$ treatments, the first row of the square would be

$$1 \quad 2 \quad 4 \quad 3$$

Then the remaining rows of the square are obtained by adding 1 to each previous row, with the understanding that if the number obtained is greater than n we then subtract n. Thus, for the 4×4 square, we obtain

$$\begin{array}{cccc} 1 & 2 & 4 & 3 \\ 2 & 3 & 1 & 4 \\ 3 & 4 & 2 & 1 \\ 4 & 1 & 3 & 2 \end{array}$$

We see that, in this Latin square, Treatment 1 follows immediately after Treatments 2, 3, and 4 one time each. Similarly, Treatment 2 follows immediately after Treatments 1, 3, and 4 one time each; Treatment 3 follows immediately after Treatments 1, 2, and 4 one time each; and Treatment 4 follows immediately after Treatments 1, 2, and 3 one time each.

If the number of treatments is odd, then two Latin squares are required to have each treatment follow immediately after every other treatment an equal number of times. For one of the two squares we follow the same sequence for the first row as when the number of treatments is even. For the other square we reverse the sequence of the first row. For 5 treatments, for example, the first row of the first square would be 1, 2, 5, 3, 4, and for

the first row of the second square we would have 4, 3, 5, 2, 1. Then, following the addition rule for each square, we obtain

Square 1					Square 2				
1	2	5	3	4	4	3	5	2	1
2	3	1	4	5	5	4	1	3	2
3	4	2	5	1	1	5	2	4	3
4	5	3	1	2	2	1	3	5	4
5	1	4	2	3	3	2	4	1	5

In these two squares each treatment follows immediately after every other treatment exactly two times.

Suppose, for example, that Treatment 4 functions as a general depressant, lowering the value of the observation by a constant for any treatment immediately following it Since every treatment follows immediately after Treatment 4 exactly twice, the means for Treatments 1, 2, and 3 will be influenced in the same way and the differences between them will not be changed. On the other hand, the means for Treatments 1, 2, and 3, relative to the mean for Treatment 4 will not be the same as they would be if Treatment 4 had no influence on the other treatments. If Treatment 4 acts nonadditively or differentially upon the treatments immediately following it, the treatment means will not be influenced in the same manner and the differences between them will not be the same as we might expect to find if each treatment mean was based upon a set of independent observations as, for example, we would have in a randomized groups design.

Analysis of Variance for Balanced Designs

The methods of analysis for Latin square designs which we have presented, in addition to assuming negligible interactions, also involve the assumption that treatment effects are constant and that there is no carry-over or residual effect from one treatment to the next. The analysis of variance for the balanced Latin squares provides an estimate of both the treatment effects and the residual effects of the immediately preceding treatment.[9] A somewhat better estimate of the residual effects can be obtained if an additional period or column is added to the Latin square which duplicates exactly the treatments of what would ordinarily be the last column or period. In this way, each treatment will be preceded equally often by every other treatment including itself. The analysis of variance for the balanced Latin square design can be found in Cochran and Cox (1957).

GRAECO-LATIN SQUARES

Suppose we have two Latin squares of order $t \times t$. To distinguish between these two squares, let the cell entries in one be represented by

[9] For another solution to the problem of residual effects, see Pearce (1957).

Latin letters and the cell entries in the other by Greek letters. If one of these squares is imposed on the other and if each Greek letter occurs once and only once with each Latin letter, then the two superimposed squares are said to form a Graeco-Latin square. The following is an example of a 4×4 Graeco-Latin square:

$$\begin{array}{cccc} \delta B & \beta D & \gamma A & \alpha C \\ \alpha A & \gamma C & \beta B & \delta D \\ \gamma D & \alpha B & \delta C & \beta A \\ \beta C & \delta A & \alpha D & \gamma B \end{array}$$

We note that each Greek letter occurs once and only once in each row and each column and that the same is true for each Latin letter. Thus the Greek letters alone meet the requirements of a Latin square as do also the Latin letters alone. Furthermore, since each Greek letter occurs once and only once with each Latin letter the two superimposed squares form a Graeco-Latin square.

To consider a possible application of a Graeco-Latin square, let the rows correspond to subjects, the columns to successive testing periods, the Latin letters to treatments, and the Greek letters to different lists of materials to be learned. Then the analysis of variance for the Graeco-Latin square would result in the following sums of squares and degrees of freedom:

Periods	$t - 1$
Subjects	$t - 1$
Treatments	$t - 1$
Lists	$t - 1$
Error	$(t - 1)(t - 3)$

where t is the order of the square. Thus, for a 4×4 Graeco-Latin square, the error sum of squares would have only 3 d.f., and for a 5×5 Graeco-Latin square the error sum of squares would have only 8 d.f. In the 6×6 and 7×7 Graeco-Latin squares, the error sums of squares have 15 and 24 d.f., respectively.

Potential applications and limitations of the Graeco-Latin square in psychological research are discussed by Archer (1952) and Grant (1948). The analysis of variance for independently replicated Graeco-Latin squares is also described by Archer (1952). Methods for constructing Graeco-Latin squares are given by Fisher and Yates (1948).

QUESTIONS AND PROBLEMS

1. Sleight (1948) used a Latin square design to study the influence of the shape of instrument dials and exposure time on legibility. Five dial shapes were used: (*H*)orizontal, (*O*)pen window, (*R*)ound, (*V*)ertical, and (*S*)emicircular. The exposure times used were .28, .20, .17, .14, and .12 seconds. In a preliminary experiment, 5 subjects were tested. The measurements reported are the number

of errors made by the subject in reading the various dials under the various exposure times. The data are given below:

Subjects	Exposure Speed in Seconds					Subjects	Exposure Speed in Seconds				
	.28	.20	.17	.14	.12		.28	.20	.17	.14	.12
1	*H*	*O*	*S*	*R*	*V*	1	3	0	4	2	6
2	*S*	*R*	*V*	*H*	*O*	2	2	2	6	1	0
3	*V*	*H*	*O*	*S*	*R*	3	10	6	1	6	0
4	*O*	*S*	*R*	*V*	*H*	4	0	4	4	12	2
5	*R*	*V*	*H*	*O*	*S*	5	3	6	8	0	7

(*a*) Analyze the original observations. Note that the means and variances for three of the dial shapes are much the same, suggesting a square root transformation. (*b*) Add .5 to each cell entry and then take the square root. Analyze the transformed data. Has the transformation tended to stabilize the variance? (*c*) Is there reason to believe that carry-over effects might be present? (*d*) Are there good arguments for or against having the longest exposure time on the first trial?

2. De Lury (1946) reports upon a Latin square design concerned with the investigation of the reactions of rabbits to four different doses of a drug. The observations reported below are in terms of milligrams of glucose per 100 cc. of blood. Four independently drawn and randomized Latin squares were used. The data are given below:

Rabbits	Days				Rabbits	Days			
	1	2	3	4		1	2	3	4
1	*C*	*A*	*B*	*D*	1	59	56	41	54
2	*B*	*D*	*C*	*A*	2	56	58	73	69
3	*A*	*C*	*D*	*B*	3	45	41	30	28
4	*D*	*B*	*A*	*C*	4	62	49	63	84
5	*A*	*B*	*D*	*C*	5	42	39	44	61
6	*C*	*A*	*B*	*D*	6	49	61	38	43
7	*B*	*D*	*C*	*A*	7	83	81	101	96
8	*D*	*C*	*A*	*B*	8	56	54	65	58
9	*B*	*D*	*A*	*C*	9	47	46	62	76
10	*A*	*C*	*B*	*D*	10	90	74	61	63
11	*C*	*B*	*D*	*A*	11	79	63	58	87
12	*D*	*A*	*C*	*B*	12	50	69	66	59
13	*D*	*A*	*C*	*B*	13	45	61	45	71
14	*C*	*B*	*D*	*A*	14	52	31	35	81
15	*A*	*D*	*B*	*C*	15	57	30	57	50
16	*B*	*C*	*A*	*D*	16	64	83	74	67

The data reported above by De Lury were made available through the courtesy of Dr. D. M. Young and the Connaught Laboratories of the University of Toronto. Analyze the results of each Latin square separately. Can we conclude that the residual mean squares are homogeneous? Recall that, as we stated

earlier, the χ^2 test is about as sensitive to nonnormality as to heterogeneity of variance. One method we might use in considering this problem would be to find the residual deviations for each square and then to examine graphically the distribution of the residuals.

3. The airplane location experiment described in the chapter was concerned not only with the ability of the subjects to locate the position of the target in terms of degrees, but also with the ability of the subjects to judge the distance of the plane from the center of the target. The latter judgments were made in terms of miles. The results are given below:

Scores of Subjects in Locating Targets on Screens of Five Different Sizes

Latin Square	Subjects	Periods					
		1	2	3	4	5	Σ
	1	19	21	25	27	22	114
	2	22	20	23	31	24	120
3 6 4 7 5	3	26	28	26	31	32	143
	4	17	20	17	14	18	86
	5	28	30	30	31	28	147
	Σ	112	119	121	134	124	610
	6	23	30	29	24	28	134
	7	24	33	28	19	32	136
4 7 5 3 6	8	29	30	31	29	28	147
	9	24	26	27	25	31	133
	10	11	18	27	18	24	98
	Σ	111	137	142	115	143	648
	11	18	16	24	24	18	100
	12	29	26	29	29	27	140
5 3 6 4 7	13	30	27	30	30	31	148
	14	25	22	28	26	30	131
	15	15	17	15	16	15	78
	Σ	117	108	126	125	121	597
	16	27	24	27	29	26	133
	17	22	14	12	18	15	81
6 4 7 5 3	18	28	30	34	31	28	151
	19	23	22	25	23	16	109
	20	29	26	28	28	26	137
	Σ	129	116	126	129	111	611
	21	29	28	21	30	24	132
	22	30	31	30	33	30	154
7 5 3 6 4	23	19	19	17	25	24	104
	24	28	20	16	20	21	105
	25	24	27	25	28	27	131
	Σ	130	125	109	136	126	626

In the analysis of variance, find the separate S's $\times$ periods sums of squares for each order. Check to see that the sum of these is equal to the sum of squares that would be obtained by subtraction.

4. Define, briefly, each of the following terms: (*a*) Latin square, (*b*) carry-over effect, (*c*) balanced Latin square.

16

THE ANALYSIS OF COVARIANCE FOR A RANDOMIZED GROUPS DESIGN

INTRODUCTION

In this chapter we shall consider the simplest form of the *analysis of covariance* as applied to a randomized groups design. In the analysis of covariance we have two observations for each subject. One of these we shall designate as a *supplementary* measure X which is not itself of experimental interest.[1] The other we shall designate as Y. The Y measures are those obtained on the dependent variable of interest after the treatments have been applied. It is the significance of the difference between the Y means for the various treatment groups that is of interest.

If the results obtained with the analysis of covariance are to have a clear interpretation, then it is essential that the X measures be uninfluenced by the particular treatments to which the subjects are assigned. One way in which we can satisfy this essential condition is to obtain the X measures prior to the application of the treatments.[2] Obviously, if the X measures are obtained prior to the application of the treatments, they cannot be affected by the treatments.

PRODUCT SUMS

In the analysis of variance for a randomized groups design with n observations for each group or treatment, we analyze the total sum of squares into the sum of squares within groups and the sum of squares

[1] A supplementary measure or observation is also referred to as a concomitant measure or observation.

[2] If the X measures are obtained *after* the application of treatments, the question of whether or not they can be considered to be unaffected by the treatments is of importance. If the X measures correspond to stable organismic variables, such as intelligence, then we might be reasonably confident that they are not influenced by the treatments.

between groups [3] We have found that we can express the deviation of a given value X_{kn} from the over-all mean $\bar{X}..$ as

$$X_{kn} - \bar{X}.. = (X_{kn} - \bar{X}_k) + (\bar{X}_{k\cdot} - \bar{X}..)$$

Similarly, a deviation of Y_{kn} from the over-all mean $\bar{Y}..$ can be expressed as

$$Y_{kn} - \bar{Y}.. = (Y_{kn} - \bar{Y}_{k\cdot}) + (\bar{Y}_{k\cdot} - \bar{Y}..)$$

Multiplying these two expressions and summing over the n observations in a single group, we obtain the product sum

$$\begin{aligned}\sum_1^n (X_{kn} - \bar{X}.)(Y_{kn} - \bar{Y}..) = &\sum_1^n (X_{kn} - \bar{X}_{k\cdot})(Y_{kn} - \bar{Y}_k) \\ &+ \sum_1^n (\bar{X}_{k\cdot} - \bar{X}..)(Y_{kn} - \bar{Y}_{k\cdot}) \\ &+ \sum_1^n (X_{kn} - \bar{X}_{k\cdot})(\bar{Y}_{k\cdot} - \bar{Y}..) \\ &+ \sum_1^n (\bar{X}_{k\cdot} - \bar{X}..)(\bar{Y}_{k\cdot} - \bar{Y}..)\end{aligned}$$

Both $(\bar{X}_{k\cdot} - \bar{X}..)$ and $(\bar{Y}_{k\cdot} - \bar{Y}..)$ are constants and we also know that the sum of the deviations of the n observations in the kth group from the mean of the group is equal to zero. Thus

$$(\bar{X}_{k\cdot} - \bar{X}..) \sum_1^n (X_{kn} - \bar{X}_{k\cdot}) = 0$$

and

$$(\bar{Y}_{k\cdot} - \bar{Y}..) \sum_1^n (Y_{kn} - \bar{Y}_{k\cdot}) = 0$$

Thus we have

$$\begin{aligned}\sum_1^n (X_{kn} - \bar{X}..)(Y_{kn} - \bar{Y}..) = &\sum_1^n (X_{kn} - \bar{X}_{k\cdot})(Y_{kn} - \bar{Y}_k) \\ &+ n(\bar{X}_{k\cdot} - \bar{X}..)(\bar{Y}_{k\cdot} - \bar{Y}..)\end{aligned}$$

for a single group. Then summing over the k groups, we have

$$\begin{aligned}\sum_1^{kn} (X_{kn} - \bar{X}..)(Y_{kn} - \bar{Y}..) = &\sum_1^{kn} (X_{kn} - \bar{X}_{k\cdot})(Y_{kn} - \bar{Y}_{k\cdot}) \\ &+ n \sum_1^k (\bar{X}_{k\cdot} - \bar{X}..)(\bar{Y}_{k\cdot} - \bar{Y}..)\end{aligned}$$

The term on the left in the above expression is called the *total product sum.* We have analyzed the total product sum into the two component

[3] Just as the analysis of variance for a randomized groups design does not require that we have equal n's in each of the treatment groups, so also this is not a necessary requirement for the analysis of covariance.

parts shown on the right. The first term on the right is the *product sum within groups* and the second term is the *product sum between groups.* We shall designate these product sums by $\sum xy_t$, $\sum xy_w$, and $\sum xy_b$, respectively. The total product sum can be obtained by finding

$$\sum xy_t = \sum_1^{kn} X_{kn}Y_{kn} - \frac{\left(\sum_1^{kn} X_{kn}\right)\left(\sum_1^{kn} Y_{kn}\right)}{kn} \tag{16.1}$$

The product sum between groups can be obtained by calculating

$$\sum xy_b = \frac{\left(\sum_1^{n} X_{1n}\right)\left(\sum_1^{n} Y_{1n}\right)}{n_1} + \frac{\left(\sum_1^{n} X_{2n}\right)\left(\sum_1^{n} Y_{2n}\right)}{n_2} + \cdots + \frac{\left(\sum_1^{n} X_{kn}\right)\left(\sum_1^{n} Y_{kn}\right)}{n_k} - \frac{\left(\sum_1^{kn} X_{kn}\right)\left(\sum_1^{kn} Y_{kn}\right)}{kn} \tag{16.2}$$

For any single group the product sum will be given by

$$\sum_1^{n} xy_k = \sum_1^{n} X_{kn}Y_{kn} - \frac{\left(\sum_1^{n} X_{kn}\right)\left(\sum_1^{n} Y_{kn}\right)}{n_k}$$

and summing over all groups, we obtain the product sum within groups. Thus

$$\sum xy_w = \sum_1^{k}\left[\sum_1^{n} X_{kn}Y_{kn} - \frac{\left(\sum_1^{n} X_{kn}\right)\left(\sum_1^{n} Y_{kn}\right)}{n_k}\right] \tag{16.3}$$

Since the sum of the product sums between groups and within groups must equal the total product sum, we can also obtain the within-groups product sum by subtraction. Thus

$$\sum xy_w = \sum xy_t - \sum xy_b \tag{16.4}$$

RELATIONSHIP BETWEEN X AND Y IN THE ABSENCE OF TREATMENT EFFECTS

In Table 16.1 we give the X and Y measures for a randomized groups design in which we have 5 subjects assigned to each of 3 treatments. Let us assume that the X measures are on the same variable as the Y measures, the only difference being that the X measure is obtained prior to the application of the treatment and the Y measure after the application of the treatment. In Table 16 1, however, the Y measures were derived in such a way that there are no significant differences between the treatment (Y) means.

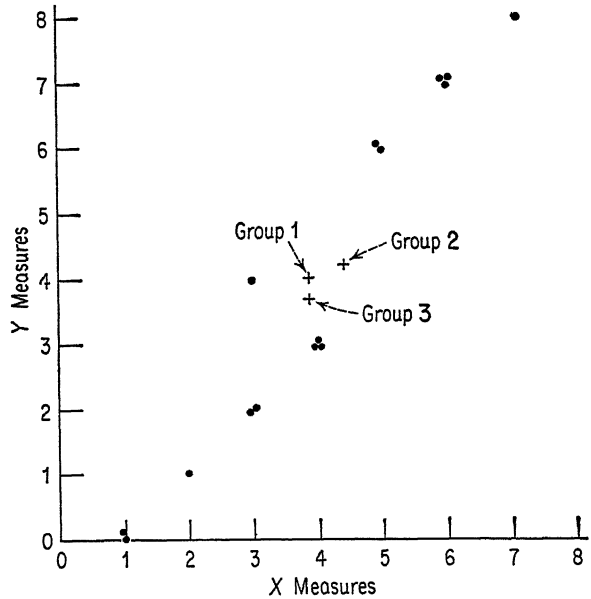


Figure 16.1 Plot of the Y measures against the X measures for the data of Table 16.1. The three + signs represent the means of the three groups on the X and Y variables.

Figure 16.1 shows the plot of the Y measures against the X measures. With fairly reliable measurements and in the absence of treatment effects, we should expect to obtain a plot of the X and Y values that is similar to that of Figure 16.1.[4] With a randomized groups design, the X values on

Table 16.1 Measures on a Supplementary Variable (X) and a Dependent Variable (Y) for a Randomized Groups Design in the Absence of Treatment Effects

	Treatment Groups					
	1		2		3	
	X	Y	X	Y	X	Y
	1	0	2	1	1	0
	6	7	3	2	4	3
	3	4	6	7	5	6
	4	3	4	3	3	2
	5	6	7	8	6	7
$\sum$	19	20	22	21	19	18

[4] The important feature of Figure 16.1 is that the trend of the points can be represented by a single regression line.

the baseline of the figure will be divided at random into 3 sets of 5 each. Since these measures are obtained prior to the application of the treatments, we should not expect to find any significant differences between the treatment groups on the X measures. In the absence of treatment effects we should not expect to find any significant differences between the groups on the Y measures either.

RELATIONSHIP BETWEEN X AND Y WHEN TREATMENT EFFECTS ARE PRESENT

We now consider the case where we do have treatment effects which are additive. Let us increase the Y values in Table 16.1 for each subject in Group 1 by 5. We leave the Y values for each subject in Group 2 unchanged. For the subjects in Group 3 we increase the Y values by 10. Making these changes in the Y measures, we obtain Table 16.2. The X

Table 16.2 Measures on a Supplementary Variable (X) and a Dependent Variable (Y) for a Randomized Groups Design with Treatment Effects Present

	Treatment Groups					
	1		2		3	
	X	Y	X	Y	X	Y
	1	5	2	1	1	10
	6	12	3	2	4	13
	3	9	6	7	5	16
	4	8	4	3	3	12
	5	11	7	8	6	17
$\sum$	19	45	22	21	19	68

values in Table 16.2 are unchanged and are the same as those shown in Table 16.1.

In Figure 16.2 we show the plot of the Y measures against the X measures for the data of Table 16.2. In this figure it is clear that for each treatment group the points cluster about relatively parallel lines at different heights. This is the sort of graph we should expect to obtain under the following conditions:

(1) We have treatment effects which are additive.
(2) In the absence of treatment effects the relationship between X and Y is positive and linear.[5]

[5] Analysis of covariance techniques can be applied to the case of nonlinear relationships. However, the formulas given in this chapter are not appropriate if the relationship between X and Y is not linear. With nonlinear relationships the methods of analysis are more complex than in the case of linear relationships.

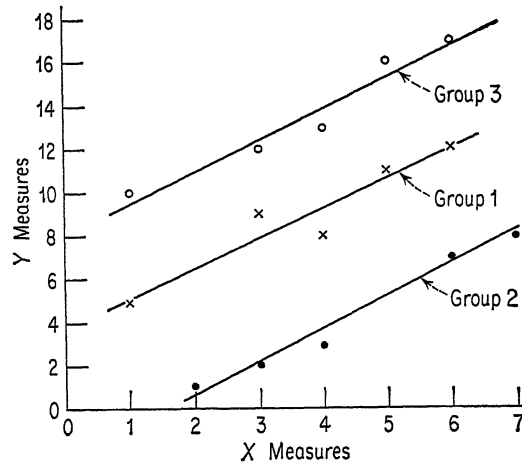


Figure 16.2 Plot of the Y measures against the X measures for each of three groups. The lines shown are the regression lines of Y on X for each of the separate groups. Original data given in Table 16.2.

(3) We have randomly assigned subjects to the treatment groups.
(4) The X measures are uninfluenced by the treatments.

SUMS OF SQUARES AND PRODUCT SUMS FOR A RANDOMIZED GROUPS DESIGN

Sums of Squares

To apply the analysis of covariance to the data of Table 16.2 we need to partition the total sum of squares on both the X and Y variable and also the total product sum. For the X measures we have

$$\sum x_t^2 = (1)^2 + (6)^2 + \cdots + (6)^2 - \frac{(60)^2}{15} = 48.00$$

and

$$\sum x_b^2 = \frac{(19)^2}{5} + \frac{(22)^2}{5} + \frac{(19)^2}{5} - \frac{(60)^2}{15} = 1.20$$

We also have

$$\sum x_1^2 = (1)^2 + (6)^2 + \cdots + (5)^2 - \frac{(19)^2}{5} = 14.80$$

$$\sum x_2^2 = (2)^2 + (3)^2 + \cdots + (7)^2 - \frac{(22)^2}{5} = 17.20$$

$$\sum x_3^2 = (1)^2 + (4)^2 + \cdots + (6)^2 - \frac{(19)^2}{5} = 14.80$$

Then the sum of squares within groups will be equal to

$$\sum x_w^2 = 14\,80 + 17.20 + 14\,80 = 46.80$$

Similarly, for the Y measures we obtain

$$\sum y_t^2 = (5)^2 + (12)^2 + \cdots + (17)^2 - \frac{(134)^2}{15} = 322.93$$

and

$$\sum y_b^2 = \frac{(45)^2}{5} + \frac{(21)^2}{5} + \frac{(68)^2}{5} - \frac{(134)^2}{15} = 220.93$$

To obtain the sum of squares within groups, we find

$$\sum y_1^2 = (5)^2 + (12)^2 + \cdots + (11)^2 - \frac{(45)^2}{5} = 30.00$$

$$\sum y_2^2 = (1)^2 + (2)^2 + \cdots + (8)^2 - \frac{(21)^2}{5} = 38.80$$

$$\sum y_3^2 = (10)^2 + (13)^2 + \cdots + (17)^2 - \frac{(68)^2}{5} = 33.20$$

Then the sum of squares within groups will be equal to

$$\sum y_w^2 = 30.00 + 38.80 + 33.20 = 102.00$$

Product Sums

We now find the total product sum. Thus

$$\sum xy_t = (1)(5) + (6)(12) + \cdots + (6)(17) - \frac{(60)(134)}{15} = 53.00$$

For the product sum between groups we have

$$\sum xy_b = \frac{(19)(45)}{5} + \frac{(22)(21)}{5} + \frac{(19)(68)}{5} - \frac{(60)(134)}{15} = -14.20$$

Table 16.3 Sums of Squares and Product Sums for the Data of Table 16.2

	$\sum x^2$	$\sum y^2$	$\sum xy$
Group 1	14.80	30.00	20.00
Group 2	17.20	38.80	25.60
Group 3	14.80	33.20	21.60
Within	46.80	102.00	67.20
Between	1.20	220.93	−14.20
Total	48.00	322.93	53.00

The product sums for each of the 3 groups are

$$\sum xy_1 = (1)(5) + (6)(12) + \cdots + (5)(11) - \frac{(19)(45)}{5} = 20.00$$

$$\sum xy_2 = (2)(1) + (3)(2) + \cdots + (7)(8) - \frac{(22)(21)}{5} = 25.60$$

$$\sum xy_3 = (1)(10) + (4)(13) + \cdots + (6)(7) - \frac{(19)(68)}{5} = 21.60$$

and the product sum within groups will be equal to

$$\sum xy_w = 20.00 + 25.60 + 21.60 = 67.20$$

The results of our calculations are summarized in Table 16.3.

VARIATION WITHIN EACH GROUP ABOUT THE REGRESSION LINE FOR THE GROUP

Consider only one of the k groups. A "best fitting" straight line can be drawn to represent the regression of the Y variable on the X variable for this group.[6] The slope of this line will be given by the regression coefficient of Y on X for which we shall use the symbol b. The regression coefficient may be written

$$b_k = \frac{\sum_1^n xy_k}{\sum_1^n x_k^2} \tag{16.5}$$

and the equation for the regression line will be

$$\tilde{y}_k = b_k x_k \tag{16.6}$$

The sum of squared deviations of the actual y_k values about the regression line with slope b_k will be equal to[7]

$$\sum_1^n (y_k - b_k x_k)^2 = \sum_1^n y_k^2 - 2b_k \sum_1^n xy_k + b_k^2 \sum_1^n x_k^2$$

[6] The line will give the "best fit" in the sense that the sum of the squared deviations of the Y values from this line will be less than from any other straight line.

[7] In this expression, we have, as before,

$$x_k = X_{kn} - \bar{X}_k. \qquad \text{and} \qquad y_k = Y_{kn} - \bar{Y}_k.$$

or the deviations of X_{kn} and Y_{kn} from the group means.

Simplifying this expression, we obtain

$$\sum_1^n (y_k - b_k x_k)^2 = \sum_1^n y_k^2 - \frac{\left(\sum_1^n xy_k\right)^2}{\sum_1^n x_k^2} \tag{16.7}$$

The sum of squares of formula (16.7) has $n - 2$ d.f., the first term on the right having $n - 1$ d.f. and second term 1 d.f.

We can obviously apply formula (16.7) to each of the k groups. Then summing these sums of squares over the k groups we obtain

$$\sum_1^k \sum_1^n (y_k - b_k x_k)^2 = \sum_1^k \sum_1^n y_k^2 - \sum_1^k \frac{\left(\sum_1^n xy_k\right)^2}{\sum_1^n x_k^2}$$

We designate this sum of squares as S_1. Since the first term on the right is the sum of squares within groups on the Y variable, we have

$$S_1 = \sum y_w^2 - \sum_1^k \frac{\left(\sum_1^n xy_k\right)^2}{\sum_1^n x_k^2} \tag{16.8}$$

S_1 is a sum of k sums of squares, each with $n - 2$ d.f. Therefore, S_1 will have $k(n - 2)$ d.f.

Taking the appropriate values from Table 16.3, we have

$$\begin{aligned} S_1 &= 102.00 - \left[\frac{(20.00)^2}{14.80} + \frac{(25.60)^2}{17.20} + \frac{(21.60)^2}{14.80}\right] \\ &= 102.00 - (27.03 + 38.10 + 31.52) \\ &= 102.00 - 96.65 \\ &= 5.35 \end{aligned}$$

with 9 d.f.

The regression coefficients for each of the 3 treatment groups, in the example under consideration, can be calculated from the data of Table 16.3. Thus we have

$$b_1 = \frac{20.00}{14.80} = 1.35$$

$$b_2 = \frac{25.60}{17.20} = 1.49$$

$$b_3 = \frac{21.60}{14.80} = 1.46$$

From the data of Table 16.3, we have

$$b_w = \frac{67.20}{46.80} = 1.44$$

The three regression lines, each with slope equal to b_w, are shown in Figure 16 3. S_2 is a measure of the variation within each group about the regression lines with common slope equal to b_w.

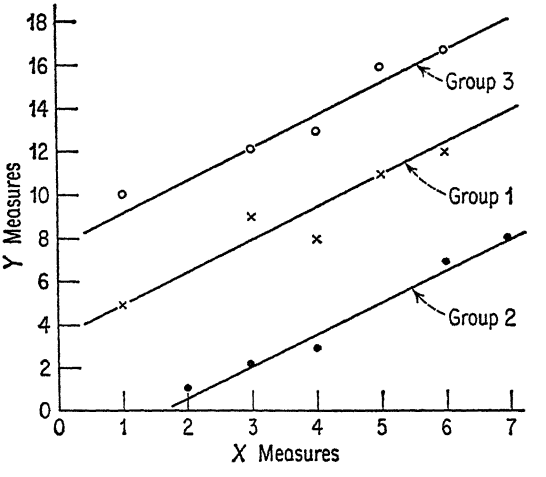


Figure 16.3 Plot of the Y measures against the X measures for each of three groups. The three regression lines each have a common slope equal to b_w. Original data given in Table 16.2.

TEST OF SIGNIFICANCE OF DIFFERENCES BETWEEN THE GROUP REGRESSION COEFFICIENTS

Now S_2 can never be smaller than S_1, since S_1 is based upon the squared deviations within each group from a regression line with slope b_k fitted separately for each group. S_2, on the other hand, is based upon the squared deviations from a regression line with the same slope b_w for each group. Thus if b_k is not equal to b_w, the sum of squared deviations for a given group with b_k as the slope of the regression line will be smaller than the sum of squared deviations obtained by using b_w as the slope of the regression line. If the b_k values do show considerable variation, then S_2 will be considerably larger than S_1.

To determine whether the regression coefficients differ significantly, we find

$$S_3 = S_2 - S_1 \tag{16.12}$$

Since S_2 has $k(n-1)-1$ d.f. and since S_1 has $k(n-2)$ d.f., we will have

$$[k(n-1)-1]-[k(n-2)]=k-1$$

degrees of freedom for S_3. For the present problem we have

$$S_3 = 5.51 - 5.35 = .16$$

with 2 d.f.

Then for the test of significance of the differences between the group regression coefficients, we have

$$F = \frac{\dfrac{S_3}{k-1}}{\dfrac{S_1}{k(n-2)}} \tag{16.13}$$

or, for the present problem,

$$F = \frac{\dfrac{.16}{2}}{\dfrac{5.35}{9}} = \frac{.08}{.59} = .14$$

with 2 and 9 d.f., and this is obviously a nonsignificant value.

If the F of formula (16.13) is significant, we would conclude that the separate regression lines are not parallel within the limits of random sampling, that is, they have significantly different slopes. This may occur because the treatment effects are not additive. In this instance, one of the transformations, described previously, applied to the Y measures may result in a scale on which the treatments are additive. It is important to stress that the application of the analysis of covariance does assume that the regression lines for the various treatment groups all can be assumed to have a common slope equal to b_w. Since, in the example under consideration, we have not obtained a significant F, we can proceed to use the analysis of covariance to test the differences between the treatment means for significance.

TEST OF SIGNIFICANCE FOR THE TREATMENT MEANS

The regression coefficient based upon the total product sum and total sum of squares on the X variable will be given by

$$b_t = \frac{\sum xy_t}{\sum x_t^2} \tag{16.14}$$

and the equation for the regression line will be

$$\tilde{y} = b_t x \tag{16.15}$$

The sum of squared deviations of the actual y values about the regression line with slope b_t will be[8]

$$\sum_{1}^{kn} (y - b_t x)^2 = \sum_{1}^{kn} y^2 + b_t^2 \sum_{1}^{kn} x^2 - 2b_t \sum_{1}^{kn} xy_t$$

and we designate this sum of squares as S_4. Simplifying the expression for S_4, we have

$$S_4 = \sum y_t^2 - \frac{(\sum xy_t)^2}{\sum x_t^2} \tag{16.16}$$

and S_4 will have $kn - 2$ d.f.

For the present example, we have

$$S_4 = 322.93 - \frac{(53\ 00)^2}{48.00} = 264.41$$

with 13 d.f.

Taking the difference between S_4 and S_2, we have an "adjusted" treatment sum of squares or

$$S_5 = S_4 - S_2 \tag{16.17}$$

The first term on the right has $kn - 2$ d.f. and the second term has $k(n - 1) - 1$ d.f. Thus, we have

$$[kn - 2] - [k(n - 1) - 1] = k - 1$$

degrees of freedom for S_5.

We may also note that

$$S_5 = \left[\sum y_t^2 - \frac{(\sum xy_t)^2}{\sum x_t^2}\right] - \left[\sum y_w^2 - \frac{(\sum xy_w)^2}{\sum x_w^2}\right]$$

or

$$S_5 = \sum y_b^2 - \left[\frac{(\sum xy_t)^2}{\sum x_t^2} - \frac{(\sum xy_w)^2}{\sum x_w^2}\right] \tag{16.18}$$

For the present problem, using formula (16.17), we have

$$S_5 = 264.41 - 5.51 = 258.90$$

[8] In this expression, we have, as before,

$$x = X_{kn} - \bar{X}.. \qquad \text{and} \qquad y = Y_{kn} - \bar{Y}..$$

or the deviations of X_{kn} and Y_{kn} from the over-all means.

and with formula (16.18)

$$S_5 = 220.93 - \left[\frac{(53.00)^2}{48.00} - \frac{(67\ 20)^2}{46.80}\right]$$
$$= 220.93 - (58.52 - 96.49)$$
$$= 258.90$$

In general, with a randomized groups design, we should not expect to have very large differences in the X means of the various groups. If the X means are identical, then $\sum x_b^2$ will be zero and we would have $\sum x_t^2 = \sum x_w^2$. Also, if the X means are identical, then $\sum xy_b$ will be zero and we would have $\sum xy_t = \sum xy_w$. Under this condition S_5 will be exactly equal to the sum of squares between groups on the Y variable.

Table 16.4 Summary of the Covariance Analysis of the Data of Table 16.2

Source of Variation	Sum of Squares	d f	Mean Square	F
S_5: Treatments	258.90	2	129 450	258.38
S_2: Error	5.51	11	.501	
S_4: Total	264.41	13		

Table 16.4 summarizes the analysis of covariance. The estimate of experimental error is the mean square for S_2. For the test of significance of the adjusted treatment mean square, S_5, we have

$$F = \frac{\dfrac{S_5}{k-1}}{\dfrac{S_2}{k(n-2)}} \tag{16.19}$$

or, for the data of Table 16.4,

$$F = \frac{\dfrac{258.90}{2}}{\dfrac{5.51}{11}} = \frac{129.450}{.501} = 258.38$$

with 2 and 13 d.f. This is a highly significant value and we conclude that the treatment means differ significantly after adjustment for the X values.[9]

NONLINEAR RELATIONSHIP BETWEEN X AND Y

In our discussion of the analysis of covariance, we have assumed that the relationship between X and Y, in the absence of treatment effects, is

[9] Kramer (1957) has described methods for extending Duncan's multiple range test to the "adjusted" means of the analysis of covariance. See also Duncan (1957).

linear This should be a fairly safe assumption, if the X and Y measures are obtained on the same variable. On the other hand, we may sometimes obtain supplementary measures on a variable that is not the same as the Y variable. If, in the absence of treatment effects, we have reason to believe that the relationship between X and Y is not linear but curvilinear, then a logarithmic transformation of the X scale may give us a new scale on which the relationship between X and Y is linear. Under this condition, the analysis of covariance would make use of the log X measures rather than the original X measures.

ANALYSIS OF DIFFERENCE MEASURES

The arithmetic of the analysis of covariance is somewhat more involved than that of the analysis of variance in that we must deal with product sums as well as with sums of squares. In some cases, the experimenter may attempt to avoid a covariance analysis by dealing with the difference between the X and Y values for each subject. We define a difference measure as

$$D_{kn} = Y_{kn} - X_{kn} \tag{16.20}$$

and the analysis of variance may be applied directly to the D measures. In Table 16.5 we show the D measures for the data of Table 16.2 which we have treated by the analysis of covariance.

Table 16.5 Difference Scores for the Data of Table 16.2

	Treatment Groups		
	1	2	3
	4	−1	9
	6	−1	9
	6	1	11
	4	−1	9
	6	1	11
$\sum$	26	−1	49

For the total sum of squares we have

$$\sum d_t^2 = (4)^2 + (6)^2 + \cdots + (11)^2 - \frac{(74)^2}{15} = 264.93$$

and the sum of squares between groups will be equal to

$$\sum d_b^2 = \frac{(26)^2}{5} + \frac{(-1)^2}{5} + \frac{(49)^2}{5} - \frac{(74)^2}{15} = 250.53$$

The sum of squares within groups can be obtained by subtraction. Thus

$$\sum d_w^2 = 264.93 - 250.53 = 14.40$$

Table 16.6 summarizes the analysis of variance of the D measures. We have $F = 104.38$ with 2 and 12 d.f., and this is a highly significant value.

Table 16.6 Summary of the Analysis of Variance of the Difference Scores of Table 16.5

Source of Variation	Sum of Squares	d f.	Mean Square	F
Treatments	250.53	2	125.26	104.38
Error	14.40	12	1.20	
Total	264.93	14		

It should be clear that, in taking $D_{kn} = Y_{kn} - X_{kn}$, we have, in essence, assumed b_w to be equal to 1.00. If b_w does not differ too greatly from 1.00, then the analysis of covariance and the analysis of variance of the D measures will give very similar results.[10] On the other hand, if b_w differs considerably from 1.00, then we can expect the estimate of experimental error based upon the analysis of variance of the D measures to be considerably larger than the estimate of experimental error obtained if we use the correct slope, b_w.

Before undertaking the analysis of variance of the D measures, it is wise to plot the Y measures against the X measures, keeping the observations for the different treatments identified in some manner. In this way it is possible to determine from the graph whether the various regression lines are parallel and also, if they are parallel, if b_w can be assumed to be approximately equal to 1.00.

A RANDOMIZED BLOCKS DESIGN AS AN ALTERNATIVE TO THE ANALYSIS OF COVARIANCE

Assume that the X measures of Table 16.2 were available for each subject prior to assigning the subjects to the treatments. Under this condition, we may consider the randomized blocks design as an alternative to the analysis of covariance. Using the X measures to form 5 blocks of 3 subjects each, we obtain Table 16.7, which also shows the random assignment of the subjects in each block to the 3 treatments. In Table 16.8 we show the Y measures for each treatment group rearranged so that the columns correspond to treatments.

[10] For a further discussion, see Cox (1957).

Table 16.7 Arrangement of Subjects in Blocks on the Basis of the X Measures of Table 16.2 and Randomization of Treatments Within Blocks

Blocks Based on X			Randomization of Treatments		
1	1	2	*A*	*C*	*B*
3	3	3	*B*	*A*	*C*
4	4	4	*B*	*C*	*A*
5	5	6	*C*	*A*	*B*
6	6	7	*A*	*C*	*B*

Table 16.8 Measures on the Dependent Variable (Y) for Subjects in a Randomized Blocks Design Based upon the Randomized Blocks of Table 16.7 and the Corresponding Y Measures of Table 16.2

	A	*B*	*C*	$\sum$
Block 1	5	1	10	16
Block 2	9	2	12	23
Block 3	8	3	13	24
Block 4	11	7	16	34
Block 5	12	8	17	37
$\sum$	45	21	68	134

Applying the analysis of variance to the data of Table 16.8 we have the following sums of squares:

$$\text{Total} = (5)^2 + (9)^2 + \cdots + (17)^2 - \frac{(134)^2}{15} = 322.93$$

$$\text{Blocks} = \frac{(16)^2}{3} + \frac{(23)^2}{3} + \cdots + \frac{(37)^2}{3} - \frac{(134)^2}{15} = 98.26$$

$$\text{Treatments} = \frac{(45)^2}{5} + \frac{(21)^2}{5} + \frac{(68)^2}{5} - \frac{(134)^2}{15} = 220.93$$

$$\text{Blocks} \times \text{treatments} = 322.93 - 98.26 - 220.93 = 3.74$$

The summary of the analysis of variance is given in Table 16.9. We note that the estimate of experimental error of the randomized blocks design does not differ too greatly from the estimate obtained with the analysis of covariance for the same data.[11]

[11] For further discussion of the relationship between randomized blocks designs and covariance analysis, see Cox (1957) and Feldt (1958).

Table 16.9 Summary of the Analysis of Variance for the Randomized Blocks Design of Table 16.8

Source of Variation	Sum of Squares	d.f	Mean Square	F
Treatments	220.93	2	110.46	235 02
Blocks	98.26	4	24.56	
Error	3.74	8	.47	
Total	322.93	14		

The randomized blocks design is a useful alternative to the analysis of covariance and it is easy to see that the arithmetic for the randomized blocks design is considerably simpler than that for the analysis of covariance. The major difference between the two methods of analysis is in terms of how we use the X measures. In the randomized blocks design they are used to form the blocks. Once the blocks are formed, we make no further use of the actual numerical values of the X measures in our analysis. With the analysis of covariance, on the other hand, we do consider the actual numerical values of the X measures in our analysis.

To use a randomized blocks design, we must have available in advance the X measures for all subjects to be used in the experiment. This is necessary so that we can arrange the subjects into blocks. We then randomize the treatments within the blocks. In some cases we may not be able to obtain the X measures prior to assigning the treatments to the subjects. For example, we may have an experimental procedure such that each subject appears in the laboratory for only one session. At this session the X measure is obtained, the treatment is applied, and the Y measure is then obtained after the application of the treatment. Under this condition we cannot use the X measures to arrange the subjects into blocks, since we do not know what these measures will be for all of the subjects involved in the experiment. We shall know the X measure for each subject at the time he is tested, but we do not know what these measures will be for those subjects who have not yet appeared in the laboratory. Under these circumstances, if we wish to use the X measures in an attempt to reduce our estimate of experimental error, we must either use the analysis of covariance or the analysis of variance of the differences measures in analyzing the data.

SEVERAL SUPPLEMENTARY MEASURES

If we have supplementary X measures on several variables for each subject, we may combine these into a single index and use this index in the analysis of covariance. A simple sum of the X measures for each subject is

not recommended, since, if the standard deviations of the X measures are different, those with the larger standard deviations will, in general, be weighted more heavily in the index than those with the smaller standard deviations. A somewhat better index can be obtained by taking[12]

$$I = \frac{X_1}{s_1} + \frac{X_2}{s_2} + \cdots + \frac{X_k}{s_k} \tag{16.21}$$

where each of the X measures is divided by its standard deviation.

A more exact method of analysis when we have several supplementary X measures is *multiple analysis of covariance.* Worked examples can be found in Snedecor (1956).

ANALYSIS OF COVARIANCE AND OTHER EXPERIMENTAL DESIGNS

The discussion of the analysis of covariance has been limited to a randomized groups design. The analysis of covariance, however, is not restricted in its application to only this design. If we use measures of one variable to group subjects into blocks and if in addition we have another supplementary measure X for each subject, we can use the analysis of covariance with a randomized blocks design. Similarly, the analysis of covariance can be used with a Latin square design. Examples and methods of analysis for these designs can be found in Snedecor (1956) and Federer (1955).

QUESTIONS AND PROBLEMS

1. A randomized groups design was used in an experiment with $n = 5$ subjects assigned to each of 3 treatments. A supplementary measure X was available for each subject. The supplementary measures and the measures on the dependent variable Y are given below.

Treatment 1		Treatment 2		Treatment 3	
X	Y	X	Y	X	Y
3	6	2	11	2	20
9	6	7	14	9	21
16	8	13	18	14	25
19	13	19	18	20	21
24	12	23	20	23	29

(*a*) Make a plot of the observations retaining the identity of the treatment groups. What are some of the things you can determine from the plot?

[12] The index described is not the best or most adequate way of combining the several X measures. See, for example, Horst (1936), Wilks (1938), and Dunnette and Hoggatt (1957).

(*b*) Express each value of X and Y as deviations from the means of the treatment groups to which they belong. Make a plot of these deviations and compare the plot with that of (*a*). How would you account for the difference between the two plots?

(*c*) By means of the analysis of variance, determine if the means for the groups differ significantly on the X measures.

(*d*) By means of the analysis of variance, determine if the means for the groups differ significantly on the Y measures.

(*e*) Determine whether the separate regression coefficients for the groups differ significantly.

(*f*) Analyze the data using the analysis of covariance. Compare the estimates of experimental error of the analysis of covariance with that obtained in (*d*).

(*g*) Note that the observations given above are arranged in such a way that across rows the X measures are fairly homogeneous. Assume that each row corresponds to a block. Analyze the Y measures assuming the design is a randomized blocks design. Compare the results of this analysis with that of (*f*).

2. In an experiment by Mowrer (1934), previously unrotated pigeons were tested for clockwise postrotational nystagmus. The rate of rotation was one revolution in 1½ seconds. An average initial score for each pigeon, based upon 2 tests, is indicated by the symbol X. The 24 pigeons were then divided into 4 groups of 6 each. Each group was then subjected to 10 daily periods of rotation under one of the experimental conditions indicated below. The rotation speed was the same as during the initial test and the rotation periods lasted 30 seconds, with a 30-second rest interval between each period. Groups 1, 2, and 3 were practiced in a clockwise direction only. For Group 4 the environment was rotated in a counterclockwise direction. At the end of 24 days of practice, each group was tested again under the same conditions as on the initial test. These records are called Y.

Group 1 Rotation of body only. Vision excluded		Group 2 Rotation of body only. Vision permitted		Group 3 Rotation of body and environment		Group 4 Rotation of environment only	
Initial X	Final Y	Initial X	Final Y	Initial X	Final Y	Initial X	Final Y
23.8	7.9	28.5	25.1	27.5	20.1	22 9	19.9
23.8	7.1	18.5	20.7	28.1	17.7	25 2	28.2
22.6	7.7	20.3	20.3	35.7	16.8	20.8	18.1
22.8	11.2	26.6	18.9	13.5	13.5	27.7	30.5
22.0	6.4	21.2	25 4	25.9	21.0	19.1	19.3
19.6	10.0	24.0	30.0	27.9	29.3	32.2	35.1

(*a*) Use the analysis of variance to determine whether the means on the X variable for the groups differ significantly. (*b*) Use the analysis of variance to determine whether the means of the groups on the Y variable differ significantly. (*c*) Analyze the results of the experiment using the analysis of covariance.

17

ANALYSIS OF VARIANCE MODELS AND EXPECTATIONS OF MEAN SQUARES

INTRODUCTION

In our earlier discussion of factorial experiments, we emphasized that, in general, we consider the levels of each factor as being selected by the experimenter because they are the ones of interest. Generalizations based upon tests of significance are therefore confined to the particular levels and combinations of levels actually investigated. The levels of a factor, in other words, are not considered to be a random selection from some larger population of levels. For example, one of the factors in an experiment may be shock with three levels or intensities. The experimenter obviously has a choice in selecting the levels or intensities of shock to be investigated. It is not likely, however, that he will decide to choose the three levels by random methods of selection from a larger population of intensities. When the treatments, or levels of factors, are *not* randomly selected, the analysis of variance model is referred to as Model I or as a *fixed effects* model.

Assume now that we have several factors and that the levels of each factor have been randomly selected from some larger populations. The analysis of variance model, in this instance, is referred to as Model II or as a *random effects* model. If the levels of some factors have been randomly selected and those of others have not, the analysis of variance model is referred to as a *mixed model.* The mixed model thus involves both fixed effects and random effects.

If all of the treatments or levels about which inferences are to be made are included in the experiment, then the treatments or levels may be regarded as fixed and Model I is appropriate for the analysis of variance. On the other hand, if generalizations and inferences about treatments or levels not included in the experiment are to be made, then the treatments or levels investigated must be randomly selected from the population of interest. When this is the case, the treatments or levels are regarded as random and Model II is appropriate for the analysis of variance.

If it can be assumed that Model II applies to a given experiment, then generalizations based upon the tests of significance are also assumed to hold for the levels or treatments in the population of interest. This assumption is warranted, of course, only if the treatments or levels actually investigated do represent a random selection from the population of interest There may be isolated instances in which Model II can be justified for a behavioral science experiment, but, in general, this model seems unrealistic. The fixed effects model and the mixed model seem to be much closer to the realities of experimental procedures in the behavioral sciences.

MODEL II: EXPECTATIONS OF MEAN SQUARES

In Table 17.1 we give the expectations of the mean squares for a factorial experiment with three factors, A, B, and C, with n replications of each treatment combination. The expectations of the mean squares are based upon Model II, the randomized effects model. The entries in each row of the table are called *components of variance.*[1] For each factor, we have

Table 17.1 Model II: Expectation of Mean Squares for a Factorial Experiment

Source	Expectation of Mean Squares
A:	$\sigma^2 + n\sigma_{abc}^2 + nb\sigma_{ac}^2 + nc\sigma_{ab}^2 + nbc\sigma_a^2$
B:	$\sigma^2 + n\sigma_{abc}^2 + na\sigma_{bc}^2 + nc\sigma_{ab}^2 + nac\sigma_b^2$
C:	$\sigma^2 + n\sigma_{abc}^2 + na\sigma_{bc}^2 + nb\sigma_{ac}^2 + nab\sigma_c^2$
$A \times B$:	$\sigma^2 + n\sigma_{abc}^2 + nc\sigma_{ab}^2$
$A \times C$:	$\sigma^2 + n\sigma_{abc}^2 + nb\sigma_{ac}^2$
$B \times C$:	$\sigma^2 + n\sigma_{abc}^2 + na\sigma_{bc}^2$
$A \times B \times C$:	$\sigma^2 + n\sigma_{abc}^2$
Within:	σ^2

assigned a letter, a, b, and c, respectively, which is used to designate a source of variation when used as a subscript and also to designate the number of levels of the source when used as a coefficient. The expectation of the mean square for a given source always includes the error variance, σ^2, and also a component of variance directly attributable to that source. Thus in Table 17.1, the expectation of the mean square for A includes σ^2 and σ_a^2, the former denoting the error variance and the latter the variance directly attributable to A. We observe in the table, however, that associated with σ_a^2 are the coefficients nbc. The coefficients of a component for a given source consist of all those letters not used as identifying subscripts for the source. Thus, the coefficients of σ_a^2 are nbc, since n, b, and c are not used

[1] Methods for obtaining the variance components or expectations of mean squares have been described by Schultz (1955) For additional discussion, see Crump (1946, 1951), Anderson and Bancroft (1952), and Villars (1951).

as identifying subscripts. The components of variance directly attributable to each source and the coefficients of the components are the last ones entered in each row of Table 17.1.

In addition to the error variance and the component directly attributable to a source, the other components of variance in the expectation of a given mean square consist of all other components whose identifying subscripts contain all of the letters necessary to describe the source under consideration. Thus for A, we also have $n\sigma_{abc}^2$, $nc\sigma_{ab}^2$, and $nb\sigma_{ac}^2$, since the subscripts for each of these components contain a and this is the only letter necessary to describe completely the source of variation under consideration. The expectations for the other mean squares are obtained in the same manner as we obtained those for A.

In Table 17.1 it should be obvious that, for the random effects model, the error (within groups) mean square provides an appropriate error term only for the $A \times B \times C$ interaction mean square. To test the two-factor interactions for significance, the appropriate error term would be the $A \times B \times C$ interaction mean square.[2] For the main effects, there is no appropriate error mean square in the table, but approximate tests for situations of this kind have been devised by Cochran (1951) and Satterthwaite (1946).

EXPECTATIONS OF MEAN SQUARES FOR A MIXED MODEL

In Table 17.2 we show the expectations of the mean squares for a mixed model, where the levels of C are regarded as random and the levels of A and B are regarded as fixed. Following Schultz (1955), for purposes of distinction we use θ^2 to represent a component due directly to a fixed effect.

We can obtain the expectations of the mean squares in Table 17.2 from those in Table 17.1. When we have fixed effects, certain components are deleted from the expectations of the mean squares of Model II. To determine which components are to be deleted, we first delete the one or more subscript letters in the components of Table 17.1 which are necessary to describe the source in which the component is listed. Then, if any one

[2] The appropriate error term for a test of a given effect is the mean square whose expectation contains all of the components that are in the expectation of the mean square for the given effect *except* the component directly attributable to the given effect. Thus to test the $A \times B \times C$ interaction effect, we have

$$F = \frac{\sigma^2 + n\sigma_{abc}^2}{\sigma^2}$$

and as a test of the $B \times C$ interaction effect, we have

$$F = \frac{\sigma^2 + n\sigma_{abc}^2 + na\sigma_{bc}^2}{\sigma^2 + n\sigma_{abc}^2}$$

Table 17.2 Expectation of Mean Squares for a Factorial Experiment with *A* and *B* Fixed and *C* Random

Source	Expectation of Mean Squares
A:	$\sigma^2 + nb\sigma_{ac}^2 + nbc\theta_a^2$
B:	$\sigma^2 + na\sigma_{bc}^2 + nac\theta_b^2$
C:	$\sigma^2 + nab\sigma_c^2$
$A \times B$:	$\sigma^2 + n\sigma_{abc}^2 + nc\theta_{ab}^2$
$A \times C$:	$\sigma^2 + nb\sigma_{ac}^2$
$B \times C$:	$\sigma^2 + na\sigma_{bc}^2$
$A \times B \times C$:	$\sigma^2 + n\sigma_{abc}^2$
Within:	σ^2

Source	Error Mean Square for Test of Significance
A:	$A \times C$
B:	$B \times C$
C:	Within
$A \times B$:	$A \times B \times C$
$A \times C$:	Within
$B \times C$:	Within
$A \times B \times C$:	Within

of the remaining subscripts specifies a fixed effect, we delete the component from the expectation of the mean square for the source. For example, consider the expectation for A in Table 17.1. For the component σ_{abc}^2 we delete the subscript a since it is necessary to describe the source under consideration. The two remaining subscripts are then b and c. Since b specifies a fixed effect, this component is deleted from the expectation of the mean square for A in Table 17.2. Similarly, with respect to σ_{ab}^2 we delete the subscript a which is necessary to describe the source under consideration. The remaining subscript is b and since this specifies a fixed effect, the component is deleted from the expectation for A in Table 17.2. For σ_{ac}^2 we again delete a which is necessary to describe the source. The remaining subscript is c and since this subscript does *not* specify a fixed effect, this component with coefficients nb remains as part of the expectation of the mean square for A in Table 17.2.[3] Following similar procedures, we

[3] When we have a fixed effect which cross-classifies with a random effect, we note that the interaction results in a component that is random in only one direction. The component, in other words, appears as part of the expectation of the mean square for the fixed effect, but not as part of the expectation of the mean square for the random effect. Consider, for example, the component σ_{bc}^2 of Table 17.2, where B is fixed and C is random The component σ_{bc}^2 appears as part of the expectation of the mean square for B, the fixed effect, but not as part of the expectation of the mean square for C, the random effect. The reason for this is that the B effect is measured over a random sampling

obtain the expectations of the other mean squares in Table 17.2 from those in Table 17.1. The appropriate error terms for the mean squares of the mixed model of Table 17.2 are given at the bottom of the table.

If the levels of C were also fixed, as well as those of A and B, then it is easy to see, in Table 17.2, that the appropriate error mean square for testing each of the main effects and interactions is that based upon replication. With A, B, and C all representing fixed effects, the appropriate analysis of variance model is Model I.

In Table 17.2, assume that the experiment is a factorial with only one replication. Then no estimate of error, σ^2, based upon replication would be available. Under this condition, no appropriate error mean square would be available for testing the C, $A \times C$, and $B \times C$ mean squares for significance, unless we can assume that the component $\sigma_{abc}{}^2$ is negligible. If this is the case, then the $A \times B \times C$ mean square will be an estimate of σ^2.

To illustrate an experiment in which the mixed model of Table 17.2 would be appropriate, let the levels of A correspond to the hours 9, 10, and 11 o'clock in the morning. Let B have two levels, corresponding to a male and female test administrator. We regard both the levels of A and B as fixed effects which do not involve random selections from larger populations. Let C, however, have 8 levels consisting of 8 senior high schools selected at *random* from a larger population of schools. In selecting schools at random, the objective is to be able to generalize the results of the fixed effects not only with respect to the schools actually investigated, but also with respect to the population from which they were selected.

In each school 60 male students are selected at random from the senior class and these students are then assigned at random in such a way that $n = 10$ students are available for each of the combinations of levels of A and B. Thus, in each school we have 6 groups of 10 students each. One group is administered a standardized test by the male and another group is administered the test by the female at 9 o'clock. Similarly, two more groups are tested at 10, and another two groups at 11 o'clock.

The expectations of the mean squares for this factorial experiment would be the same as those given in Table 17.2. A (hour of testing) and B (sex of the test administrator) represent fixed effects. The appropriate error mean square for A is the $A \times C$ interaction and the appropriate error mean square for B is the $B \times C$ interaction. If the A mean square is significant, we would conclude that the means for the hours of testing investigated, 9, 10, and 11 o'clock, differ significantly. Since this effect is tested

---

of C and the B effect may be expected to show random variation for different random samples of C. Thus $\sigma_{bc}{}^2$ appears as a component of the expectation of the mean square for B. On the other hand, C is measured over fixed, constant levels of B. Thus there is no uncertainty associated with the C effect as a result of being measured over random samples of B, since B is not random.

over a random sample of schools, we may assume that the differences would hold also for the population of schools. Similarly, if the B mean square is significant, when tested against the $B \times C$ interaction, we would conclude that the mean scores for students tested by the male and female administrators differ significantly. Since this effect is tested over a random sample of schools, we may assume that the difference would also be true of the population of schools. Thus, by randomly selecting the levels of C (schools) we are in a position where we can generalize about a significant fixed effect not only with respect to the schools involved in the experiment, but also with respect to the larger population of schools from which we have a random selection.

EXPECTATIONS OF MEAN SQUARES IN A RANDOMIZED BLOCKS DESIGN

Consider now a randomized blocks design such that within each block we have not a single subject for each treatment but n subjects. For example, suppose we have 5 blocks of 8 subjects each such that within each block the subjects are relatively homogeneous on a variable which we believe to be relevant to the measures to be obtained after the application of treatments. Assume that we have only 2 treatments so that within each block we can randomly assign $n = 4$ subjects to each treatment. For a given treatment in a given block, it will be possible to obtain a sum of squares based upon the variation of the 4 subjects assigned to the treatment. This sum of squares will have $n - 1 = 3$ d.f. and the pooled sum of all of these sums of squares will have 30 d.f. We shall refer to this pooled sum of squares as the sum of squares within cells. The analysis of variance for this experiment would result in the following analysis:

Source	d.f.
Treatments	1
Blocks	4
Blocks × treatments	4
Within cells	30
Total	39

Table 17.3 shows the expectations of the mean squares for Model II (random effects), Model I (fixed effects), and the mixed model, where blocks are assumed to be random and treatments fixed. It is clear, for Model II, that the appropriate error mean square for testing the significance of the treatment mean square is the blocks × treatments interaction. This is also true for the mixed model. For the fixed effects model, where both blocks and treatments are regarded as fixed, the appropriate error mean square for treatments is the within-cells mean square.

Table 17.3 Expectation of Mean Squares for a Randomized Blocks Design

Model II· Blocks and Treatments Random

Source	Expectation of Mean Square
Treatments	$\sigma^2 + n\sigma_{bt}^2 + nb\sigma_t^2$
Blocks	$\sigma^2 + n\sigma_{bt}^2 + nt\sigma_b^2$
Blocks × treatments	$\sigma^2 + n\sigma_{bt}^2$
Within	σ^2

Model I· Blocks and Treatments Fixed

Source	Expectation of Mean Square
Treatments	$\sigma^2 + nb\theta_t^2$
Blocks	$\sigma^2 + nt\theta_b^2$
Blocks × treatments	$\sigma^2 + n\theta_{bt}^2$
Within	σ^2

Mixed Model: Blocks Random and Treatments Fixed

Source	Expectation of Mean Square
Treatments	$\sigma^2 + n\sigma_{bt}^2 + nb\theta_t^2$
Blocks	$\sigma^2 + nt\sigma_b^2$
Blocks × treatments	$\sigma^2 + n\sigma_{bt}^2$
Within	σ^2

In the randomized blocks design, as we have presented it earlier, we ordinarily do not have an estimate of experimental error based upon a within-cells sum of squares. Without replication of the treatments within each block, we have only the blocks × treatments mean square as an estimate of experimental error. In general, we shall not be able to regard the treatments as randomly selected and therefore we shall have to assume that either the mixed model or the fixed effects model is applicable. For the mixed model, with blocks random, the blocks × treatments mean square provides an appropriate error term for treatments. With the fixed effects model, the blocks × treatments mean square provides an appropriate error term for the treatment mean square only if we can assume that the blocks × treatments component is negligible, that is, only if θ_{bt}^2 is negligible. If θ_{bt}^2 is negligible, then the blocks × treatments mean square provides an estimate of σ^2.

As another illustration of a mixed model, we consider an experiment described by Schroeder (1945) in which various factors influencing archery performance were investigated.[4] Eleven women shot at targets at ranges of 30, 40, and 50 yards. Each subject shot at each range on each of 6 days

[4] Only one aspect of this research is considered here. For a different analysis of the same research, see Walker and Lev (1953).

with a different order of shooting on each day. The 6 possible orders are as follows:

30	40	50
30	50	40
50	30	40
50	40	30
40	50	30
40	30	50

Nine scores were obtained for each subject, each score corresponding to the sum for a given range in a given position. Thus a given subject would have a score for the 30 yard range in the first position, in the second position, and in the third position, and similarly for the 40 yard and 50 yard ranges. The total number of observations thus consisted of 99.

We let $s = 11$ correspond to the number of archers or subjects, $r = 3$ correspond to the number of ranges, and $p = 3$ correspond to the number of positions.[5] Assuming that subjects are random and that the ranges and positions correspond to fixed effects, then the expectations of the mean squares are as shown in Table 17.4.

The appropriate error terms for the various tests of significance are shown at the bottom of Table 17.4. In general, as Schultz (1955) suggests, we ordinarily regard the three separate error estimates, subjects × ranges, subjects × positions, and subjects × ranges × positions, in designs of this

Table 17.4 Expectation of Mean Squares for a Factorial Experiment in a Randomized Blocks Design with Ranges and Positions Fixed and Blocks Random

Source	d.f.	Expectation of Mean Square
Ranges	2	$\sigma^2 + p\sigma_{sr}^2 + sp\theta_r^2$
Positions	2	$\sigma^2 + r\sigma_{sp}^2 + sr\theta_p^2$
$R \times P$	4	$\sigma^2 + \sigma_{spr}^2 + s\theta_{rp}^2$
Subjects	10	$\sigma^2 + rp\sigma_s^2$
$S \times R$	20	$\sigma^2 + p\sigma_{sr}^2$
$S \times P$	20	$\sigma^2 + r\sigma_{sp}^2$
$S \times R \times P$	40	$\sigma^2 + \sigma_{spr}^2$

Source	Error Mean Square for Test of Significance
Ranges	Subjects × Ranges
Positions	Subjects × Positions
Ranges × Positions	Subjects × Ranges × Positions

[5] It should be clear that this experiment is a factorial experiment with $pr = (3)(3) = 9$ treatment combinations arranged in a randomized blocks design with $s = 11$ blocks.

kind, as homogeneous and therefore the corresponding error sums of squares and degrees of freedom would be combined to obtain a pooled subjects × ranges-positions mean square with 80 d.f.[6] Then the sum of squares for ranges-positions treatment combinations (8 d.f.) can be broken down into the orthogonal set of comparisons: ranges (2 d.f.), positions (2 d.f.), and ranges × positions (4 d.f.) for testing against the single error term with 80 d.f. This single error term, as we have shown earlier, would consist of the pooled sums of squares for subjects × ranges (20 d.f), subjects × positions (20 d.f.), and subjects × ranges × positions (40 d.f.).

EXPECTATIONS OF MEAN SQUARES IN SPLIT-PLOT DESIGNS

Let us suppose we have an experiment in which three factors are of interest, A, B, and C, and that each is at two levels. We have 8 subjects and they are divided at random in such a way that 4 are assigned to A_1 and 4 to A_2. Each subject is to be tested under all 4 BC combinations and these are randomized over 4 periods of testing. We assume that the BC treatment combinations are such that there are no carry-over effects. The layout of the experimental design is shown in Table 17.5.

As we have previously pointed out, when we have the levels of one factor randomly assigned to blocks (subjects in this instance) and the levels

Table 17.5 Layout of a Split-Plot Design with Subjects Assigned at Random to Levels of A and BC Combinations Randomized for Each Subject in Each Level of A

	Subjects	Randomization of BC Combinations			
A_1	1	B_2C_1	B_1C_2	B_1C_1	B_2C_2
	2	B_2C_2	B_2C_1	B_1C_2	B_1C_1
	3	B_1C_2	B_1C_1	B_2C_1	B_2C_2
	4	B_2C_2	B_2C_1	B_1C_2	B_1C_1
A_2	1	B_2C_1	B_1C_1	B_1C_2	B_2C_2
	2	B_2C_2	B_1C_2	B_1C_1	B_2C_1
	3	B_2C_2	B_2C_1	B_1C_2	B_1C_1
	4	B_1C_1	B_1C_2	B_2C_1	B_2C_2

of the remaining factors or combinations of the levels are assigned at random within each block, the design is called a split-plot design. We let $s = 4$ be the number of subjects in each level of A, and we have $a = 2$ levels of A. We also have $b = 2$ levels of B and $c = 2$ levels of C. To determine the expectations of the mean squares for designs of this kind,

[6] See, for example, the earlier discussion of the error mean square given by formula (13.5).

the subscript or subscripts which serve to indicate the position in the hierarchy in which a component arises are enclosed in parentheses and those which describe the source are left outside. For example, for the component corresponding to between subjects in a given level of A, we would write $\sigma_{s(a)}^2$. The subscript s describes the source, between subjects, while (a) serves to indicate that the component arises within each level of A. The subscripts describing the source, those outside the parentheses, are called "essential" by Schultz (1955). In determining the expectations of the mean squares for designs of the kind described, we follow the same rules presented earlier except that now we delete a component from the expectation of a mean square, after deleting the components describing the source, only if any of the remaining "essential" subscripts specifies a fixed effect. In the experiment being discussed, we regard subjects as random and A, B, and C as fixed. Thus, we obtain the expectations of the mean squares shown in Table 17.6.

Table 17.6 Expectation of Mean Squares for the Split-Plot Design of Table 17.5

Source	d f.	Expectation of Mean Square
A	1	$\sigma^2 + bc\sigma_{s(a)}^2 + sbc\theta_a^2$
$S's(A)$	6	$\sigma^2 + bc\sigma_{s(a)}^2$
B	1	$\sigma^2 + bc\sigma_{s(a)b}^2 + sac\theta_b^2$
$A \times B$	1	$\sigma^2 + bc\sigma_{s(a)b}^2 + sc\theta_{ab}^2$
$S's(A) \times B$	6	$\sigma^2 + bc\sigma_{s(a)b}^2$
C	1	$\sigma^2 + b\sigma_{s(a)c}^2 + sa\theta_c^2$
$A \times C$	1	$\sigma^2 + b\sigma_{s(a)c}^2 + sb\theta_{ac}^2$
$S's(A) \times C$	6	$\sigma^2 + b\sigma_{s(a)c}^2$
$B \times C$	1	$\sigma^2 + \sigma_{s(a)bc}^2 + sa\theta_{bc}^2$
$A \times B \times C$	1	$\sigma^2 + \sigma_{s(a)bc}^2 + s\theta_{abc}^2$
$S's(A) \times B \times C$	6	$\sigma^2 + \sigma_{s(a)bc}^2$

One may assume, as we ordinarily do, that the various estimates of error involving interactions of subjects and combinations of BC are homogeneous. Then pooling these sums of squares and degrees of freedom, we obtained the pooled subjects $\times$ combinations of BC mean square with 18 d.f. The BC treatment combinations (3 d.f.) can then be analyzed into the orthogonal comparisons B, C, and $B \times C$, each with 1 d.f., and each can then be tested for significance with the single error term. Similarly, the interactions of A with the BC combinations (3 d f.) can be analyzed into the $A \times B$, $A \times C$, and $A \times B \times C$ orthogonal comparisons, each with 1 d.f., and the appropriate error term for these comparisons will also be the pooled subjects $\times$ BC combinations with 18 d.f. For testing the significance

of the A mean square, we use the pooled between subjects, in each level of A, mean square with 6 d.f. as the error term.

Thus, under the assumptions made above, we would have the following sums of squares and degrees of freedom:

Source	d.f.
A	1
Error (a)	6
B	1
C	1
$B \times C$	1
$A \times B$	1
$A \times C$	1
$A \times B \times C$	1
Error (b)	18
Total	31

where error (a) is based upon the pooled sum of squares between subjects in each level of A and error (b) is based upon the pooled subjects $\times$ BC treatment combinations.

EXPECTATIONS OF MEAN SQUARES IN THE LATIN SQUARE DESIGN

Suppose we have a Latin square design with t rows, t columns, and t treatments. If we can assume that *all* interactions (rows $\times$ columns $\times$ treatments, rows $\times$ columns, rows $\times$ treatments, and columns $\times$ treatments) are negligible, then with Model I, we have the expectations of the mean squares given in Table 17.7. Thus with a fixed effects model and with all interactions zero, the treatment mean square divided by the error mean

Table 17.7 Expectation of Mean Squares for a Latin Square Design with Rows, Columns, and Treatments Fixed and with All Interactions Zero

Source	Expectation of Mean Square
R	$\sigma^2 + t\theta_r^2$
C	$\sigma^2 + t\theta_c^2$
T	$\sigma^2 + t\theta_t^2$
E	σ^2

square provides a test of significance of the treatment effects.

Consider now what happens if the interactions are not zero.[7] If the rows $\times$ columns $\times$ treatments interaction is not zero, then it will be a component of the expectations of each of the mean squares. If the rows $\times$

[7] See, for example, the article by Wilk and Kempthorne (1957).

columns interaction is not zero, it will be a component of the expectation of both the error mean square *and* the treatment mean square. If the rows × treatments interaction is not zero, it will be a component of both the error mean square *and* the column mean square. If the columns × treatments interaction is not zero, it will be a component of both the error mean square *and* the row mean square. Thus, if these interactions are not zero, we have the expectations of the mean squares shown in Table 17.8.

Table 17.8 Expectation of Mean Squares for a Latin Square Design with Rows, Columns, and Treatments Fixed and with Interactions Present

Source	Expectation of Mean Square
R	$\sigma^2 + \theta_{rct}^2 + \theta_{ct}^2 + t\theta_r^2$
C	$\sigma^2 + \theta_{rct}^2 + \theta_{rt}^2 + t\theta_c^2$
T	$\sigma^2 + \theta_{rct}^2 + \theta_{rc}^2 + t\theta_t^2$
E	$\sigma^2 + \theta_{rct}^2 + \theta_{rc}^2 + \theta_{rt}^2 + \theta_{ct}^2$

It should be clear that, with the fixed effects model, the test of significance of the treatment mean square will be biased unless we can assume that both the rows × treatments and the columns × treatments interactions are zero. If either of these interactions is not zero, then we shall be testing the treatment mean square with an inflated error mean square. Thus, if we fail to obtain a significant treatment mean square with a Latin square design, this may be because of the presence of either a rows × treatments or a columns × treatments interaction.

Since the fixed effects model gives the expectations of the mean squares with the minimum number of components for each source, if we now regard some of the sources as random, we will introduce additional components into the expectations of the mean squares for some of the sources. In general, we shall not have a random selection of columns or of treatments. If rows correspond to subjects, then it is not unusual to regard the rows as representing a random selection of subjects. If rows are random, then it should be clear that σ_{rt}^2 will be a component of the treatment mean square since, if we delete t, the subscript necessary to describe the source, the

Table 17.9 Expectation of Mean Squares for a Latin Square Design with Rows Random and Columns and Treatments Fixed and with Interactions Present

Source	Expectation of Mean Squares
R	$\sigma^2 + \sigma_{rct}^2 + \theta_{ct}^2 + t\sigma_r^2$
C	$\sigma^2 + \sigma_{rct}^2 + \sigma_{rt}^2 + t\theta_c^2$
T	$\sigma^2 + \sigma_{rct}^2 + \sigma_{rc}^2 + \sigma_{rt}^2 + t\theta_t^2$
E	$\sigma^2 + \sigma_{rct}^2 + \sigma_{rc}^2 + \sigma_{rt}^2 + \theta_{ct}^2$

remaining subscript is r and this corresponds to a random effect. Thus, for the mixed model, with rows random and columns and treatments fixed, we have the expectations of the mean squares shown in Table 17.9.

For the mixed model of Table 17.9, it should be clear that if we are to have an unbiased test of significance of the treatment mean square, the columns $\times$ treatments interaction must be zero. If the columns $\times$ treatments interaction is not zero and we fail to obtain a significant treatment mean square, then this may be because the columns $\times$ treatments interaction has resulted in our obtaining an inflated estimate of experimental error.

QUESTIONS AND PROBLEMS

1. Given a factorial experiment with a levels of A, b levels of B, and c levels of C, with n replications. (*a*) Write out the expectations of the mean squares assuming a random effects model. (*b*) Write out the expectations of the mean squares assuming a mixed model with A random and B and C fixed. (*c*) Write out the expectations of the mean squares for the mixed model with A and B fixed and C random. (*d*) Write out the expectations of the mean squares with A, B, and C fixed.

2. Assume we have a factorial experiment with three factors, A, B, and C, each at two levels. The design is a randomized blocks design with 10 blocks of subjects with 8 subjects in each block. (*a*) Assuming the three factors are fixed, write out the expectations of the mean squares and give the degrees of freedom associated with each. (*b*) What pooled sum of squares would ordinarily be used as an estimate of experimental error and how many degrees of freedom would this sum of squares have?

3. Assume we have 10 blocks of subjects with each block containing 4 subjects. Factor A with two levels is assigned at random to the blocks so that for each level there are 5 blocks. Factors B and C are each at two levels and the BC treatment combinations are assigned at random within each block. The design is a split-plot design. (*a*) Write out the expectations of the mean squares assuming that A, B, and C are fixed effects. Show the degrees of freedom associated with each mean square. (*b*) What would be the error term for testing the A effect? (*c*) What sums of squares would ordinarily be pooled to provide an error term for testing the BC treatment combinations and the $A \times BC$ interactions?

4. Describe a factorial experiment with three factors A, B, and C, in which the levels of A can be considered as representing a random selection from a larger population.

5. Describe a factorial experiment with three factors, A, B, and C, in which the levels of A and the levels of B can be considered as representing random selections from larger populations.

6. Discuss the difference between Model I, Model II, and the mixed model of the analysis of variance.

REFERENCES

ANDERSON, R. L , and T. A. BANCROFT. *Statistical theory in research* New York: McGraw-Hill, 1952

ARCHER, E. J Some Greco-Latin analysis of variance designs for learning studies. *Psychol. Bull.*, 1952, **49**, 521–537.

BARTLETT, M. S. Recent work on the analysis of variance *J. R. statist. Soc. Suppl.*, 1934, **1**, 252–255.

——— The effect of non-normality on the *t* distribution. *Proc. Camb. phil. Soc.*, 1935, **31**, 223–231.

——— Square-root transformation in analysis of variance. *J. R. statist. Soc. Suppl.*, 1936, **3**, 68–78.

——— Some examples of statistical methods of research in agriculture and applied biology. *J. R. statist. Soc. Suppl.*, 1937, **4**, 137–170.

——— The use of transformations. *Biometrics*, 1947, **3**, 39–52.

BLISS, C. I. The analysis of field experimental data expressed in percentages. *Plant Protection*, 1937, Leningrad, No. 12, 67–77.

———, and C. L. ROSE. The assay of parathyroid extract from the serum calcium of dogs *Amer. J. Hyg.*, 1940, **31**, 79–98.

BOX, G. E. P. Non-normality and tests on variances. *Biometrika*, 1953, **40**, 318–335.

——— Some theorems on quadratic forms applied in the study of analysis of variance problems. I. Effects of inequality of variance in the one-way classification. *Ann. math. Statist.*, 1954*a*, **25**, 290–302.

——— Some theorems on quadratic forms applied in the study of analysis of variance problems. II. Effects of inequality of variance and of correlation between errors in the two-way classification. *Ann. math. Statist.*, 1954*b*, **25**, 484–498.

BROWN, C. W., and E. E. GHISELLI. *Scientific method in psychology*. New York: McGraw-Hill, 1955.

BURKE, C. J. Computation of the level of significance in the *F*-test. *Psychol. Bull.*, 1951, **48**, 392–397.

BURT, C , AND W. L. GREGORY Scientific method in psychology. II. *Brit. J. statist. psychol.*, 1958, **11**, 105–128.

CHILD, I. L. Children's preference for goals easy or difficult to obtain. *Psychol. Monogr.*, 1946, No. 280.

CLARK, M., and D. A. WORCESTER. A comparison of the results obtained from the teaching of shorthand by the word unit method and the sentence method. *J. educ. Psychol*, 1932, **23**, 122–131.

COCHRAN, W. G. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, 1947, **3**, 22–38.

——— The comparison of percentages in matched samples. *Biometrika*, 1950, **37**, 256–266.

——— Testing a linear relation among variances. *Biometrics*, 1951, **7**, 17–32.

COCHRAN, W. G. Some methods for strengthening the common χ^2 tests. *Biometrics*, 1954, **10**, 417–451.

——— Analysis of covariance: Its nature and uses. *Biometrics*, 1957, **13**, 261–281.

——— and GERTRUDE M. COX. *Experimental designs.* (2nd ed.) New York: Wiley, 1957.

COX, D. R. The use of a concomitant variable in selecting an experimental design. *Biometrika*, 1957, **44**, 150–158.

——— *Planning of experiments.* New York: Wiley, 1958.

CRESPI, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.

CRUMP, S. L. The estimation of variance components in analysis of variance. *Biometrics*, 1946, **2**, 7–11.

——— The present status of variance component analysis. *Biometrics*, 1951, **7**, 1–16.

CURTISS, J. H. On transformations used in the analysis of variance. *Ann. math Statist.*, 1945, **14**, 107–122.

DE LURY, D. B. The analysis of Latin squares when some observations are missing. *J. Amer. statist. Ass.*, 1946, **41**, 370–389.

DUNCAN, D. B Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

——— Multiple range tests for correlated and heteroscedastic means. *Biometrics*, 1957, **13**, 164–176.

DUNNETT, C. W. A mult ple comparison procedure for comparing several treatments with a control. *J. Amer. statist. Ass.*, 1955, **50**, 1096–1121.

DUNNETTE, M. D., and A. C. HOGGATT. Deriving a composite score from several measures of the same attribute. *Educ. psychol. Measmt*, 1957, **17**, 423–434.

EDWARDS, A. L. Note on the "correction for continuity" in testing the significance of the difference between correlated proportions. *Psychometrika*, 1948, **13**, 185–187.

——— On "the use and misuse of the chi-square test"—the case of the 2×2 contingency table. *Psychol. Bull.*, 1950*a*, **47**, 341–346.

——— Homogeneity of variance and the Latin square design *Psychol. Bull.*, 1950*b*, **47**, 118–129.

——— Balanced Latin-square designs in psychological research. *Amer. J. Psychol.*, 1951, **64**, 598–603.

——— *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954*a*.

——— Experiments: Their planning and execution. In G. Lindzey (Ed.) *Handbook of social psychology.* Cambridge: Addison-Wesley, 1954*b*, 259–288.

——— *Statistical analysis.* (Rev. ed.) New York: Rinehart, 1958.

———, and P. HORST. The calculation of sums of squares for interactions in the analysis of variance. *Psychometrika*, 1950, **15**, 17–24.

EISENHART, C. The assumptions underlying the analysis of variance. *Biometrics*, 1947, **3**, 1–21.

FEDERER, W. T. *Experimental design.* New York: Macmillan, 1955.

FELDT, L. S. A comparison of the precision of three experimental designs employing a concomitant variable. *Psychometrika*, 1958, **23**, 335–354.

FESTINGER, L., and D. KATZ. *Research methods in the behavioral sciences.* New York: Dryden, 1953.

FINNEY, D. J. The Fisher-Yates test of significance in 2×2 contingency tables. *Biometrika*, 1948, **35**, 145–156.

FISHER, R. A. On the "probable error" of a coefficient of correlation. *Metron*, 1921, **1**, Part 4, 1–32.

——— Discussion on Dr. Wishart's paper. *J. R. statist. Soc. Suppl.*, 1934, **1**, 51–53.

——— *Statistical methods for research workers.* (6th ed.) Edinburgh: Oliver & Boyd, 1936.

——— *The design of experiments.* (3rd ed.) Edinburgh: Oliver & Boyd, 1942.

FISHER, R. A. *Statistical methods and scientific inference.* New York: Hafner, 1956.

———, and F. YATES *Statistical tables for biological, agricultural and medical research.* (3rd ed.) Edinburgh: Oliver & Boyd, 1948.

FREEMAN, M F., and J. W. TUKEY. Transformations related to the angular and the square root. *Ann Math. Statist.*, 1950, **21**, 607–611.

FRENCH, ELIZABETH G., and F H. THOMAS. The relation of achievement motivation to problem-solving effectiveness. *J. abnorm. soc. Psychol.*, 1958, **56**, 45–48.

GAITO, J. The single Latin square design in psychological research *Psychometrika*, 1958, **23**, 369–378.

GEISSER, S. A method for testing for treatment effects in the presence of learning. *Biometrics*, 1959, **15**, 389–395.

GLANVILLE, A. D., G. L. KREEZER, and K. M. DALLENBACH The effect of type-size on accuracy of apprehension and speed of localizing words. *Amer. J. Psychol.*, 1946, **59**, 220–235.

GOODMAN, L A., and W. H. KRUSKAL. Measures of association for cross classifications. *J. Amer. statist. Ass.*, 1954, **49**, 723–764.

———, and ——— Measures of association for cross classifications. II: Further discussion and references. *J. Amer. statist. Ass.*, 1959, **54**, 123–163.

GOURLAY, N. Covariance analysis and its applications in psychological research. *Brit. J. statist. Psychol.*, 1953, **6**, 25–34.

——— *F*-test bias for experimental designs in educational research. *Psychometrika*, 1955*a*, **20**, 227–248.

——— *F*-test bias for experimental designs of the Latin square type. *Psychometrika*, 1955*b*, **20**, 273–287.

GRAHAM, F. K., and B. S. KENDALL. Performance of brain-damaged cases on a memory for designs test. *J. abnorm. soc. Psychol.*, 1946, **41**, 303–314.

GRANDAGE, A. Orthogonal coefficients for unequal intervals. *Biometrics*, 1958, **14**, 287–289.

GRANT, D. A. The Latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1948, **45**, 427–442.

——— Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

———, and A. S. PATEL. Effect of an electric shock stimulus upon the conceptual behavior of "anxious" and "non-anxious" subjects. *J. gen. Psychol.*, 1957, **57**, 247–256.

GUILFORD, J. P. *Psychometric methods.* (2nd ed.) New York: McGraw-Hill, 1954.

HAGGARD, E. A. Experimental studies in affective processes: II. On the quantification and evaluation of "measured" changes in skin resistance. *J. exp. Psychol.*, 1945, **35**, 46–56.

HARTMAN, G. Application of individual taste differences towards phenylthio-carbamide in genetic investigations. *Ann. Eugen.*, Camb., 1939, **9**, 123–135.

HELLMAN, M. A study of some etiological factors of malocclusion. *Dent. Cosmos*, 1914, **56**, 1017–1032.

HORST, P. Obtaining a composite measure from a number of different measures of the same attribute. *Psychometrika*, 1936, **1**, 53–60.

HOTELLING, H. The selection of variates for use in prediction, with some comments on the general problem of nuisance parameters. *Ann. Math. Statist.*, 1940, **11**, 271–283.

JAHODA, MARIE, S COOK, and M. DEUTSCH. *Research methods in social relations.* (Rev. ed.) C. Selltiz (Ed.) New York. Holt, 1959.

KEMPTHORNE, O. *The design and analysis of experiments.* New York: Wiley, 1952.

——— The randomization theory of experimental inference. *J. Amer. statist. Ass.*, 1955, **50**, 946–967.

KENDALL, M. G., and B. B. SMITH. Randomness and random sampling numbers. *J. R. statist. Soc.*, 1938, **101**, 147–166.

KENDALL, M. G., and B. B. SMITH. Second paper on random sampling numbers, *J. R. statist. Soc. Suppl.*, 1939, **6**, 51–61.

KENDLER, H. H. Drive interaction: I. Learning as a function of the simultaneous presence of the hunger and thirst drives. *J. exp Psychol.*, 1945, **35**, 96–109.

KISH, L. Some statistical problems in research design. *Amer. sociol. Rev.*, 1959, **24**, 328–338.

KOGAN, L. S. Variance designs in psychological research. *Psychol. Bull.*, 1953, **50**, 1–40.

KRAMER, C. Y Extension of multiple range tests to group means with unequal numbers of replications *Biometrics*, 1956, **12**, 307–310.

——— Extension of multiple range tests to group correlated adjusted means. *Biometrics*, 1957, **13**, 13–18.

KUENNE, M. R Experimental investigation of the relation of language to transposition behavior in young children. *J. exp. Psychol.*, 1946, **36**, 471–490.

LATSCHA, R. Tests of significance in a 2×2 contingency table. *Biometrika*, 1953, **40**, 74–86.

LEWIS, H. B. An experimental study of the role of the ego in work. I. The role of the ego in cooperative work. *J. exp. Psychol.*, 1944, **34**, 113–126.

———, and M. FRANKLIN. An experimental study of the role of the ego in work. II. The significance of task-orientation in work. *J. exp. Psychol.*, 1944, **34**, 195–215.

LINDZEY, G. (Ed.) *Handbook of social psychology.* (Vol. 1) Cambridge: Addison-Wesley, 1954.

MAIER, N. R. F. Reasoning in humans. III. The mechanisms of equivalent stimuli and of reasoning. *J. exp. Psychol.*, 1945, **35**, 349–360.

MAINLAND, D., L. HERRERA, and M. I SUTCLIFFE. *Tables for use with binomial samples.* New York: New York University College of Medicine, Department of Medical Statistics, 1956.

MC HUGH, R. B , and P. C. APOSTOLAKOS. Methodology for the comparison of clinical with actuarial predictions. *Psychol. Bull*, 1959, **56**, 301–308.

MC NEMAR, Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 1947, **12**, 153–157.

——— On the use of Latin squares in psychology. *Psychol. Bull.*, 1951, **48**, 398–401.

MERRINGTON, MAXINE, and CATHERINE M. THOMPSON. Table of percentage points of the inverted beta (*F*) distribution. *Biometrika*, 1943, **33**, 73–88.

MERRITT, C. B., and R. G. FOWLER. The pecuniary honesty of the public at large. *J. abnorm. soc. Psychol.*, 1948, **43**, 90–93.

MOORE, K. The effect of controlled temperature changes on the behavior of the white rat. *J. exp. Psychol.*, 1944, **34**, 70–79.

MOORE, P G., and J. W. TUKEY. Queries *Biometrics*, 1954, **10**, 562–568.

MORGAN, C. T. The statistical treatment of hoarding data. *J. comp. Psychol*, 1945, **38**, 247–256.

——— *Introduction to psychology.* New York: McGraw-Hill, 1956.

MORGAN, J. J. B. Value of wrong responses in inductive reasoning. *J. exp. Psychol.*, 1945, **35**, 141–146.

MOSTELLER, F. and R. R. BUSH. Selected quantitative techniques In G. Lindzey (Ed) *Handbook of social psychology.* Cambridge· Addison-Wesley, 1954, 289–334.

MOWRER, O. H. The modification of vestibular nystagmus by means of repeated elicitation. *Comp. Psychol. Monogr.*, 1934, No. 5.

MUELLER, C. G. Numerical transformations in the analysis of experimental data. *Psychol. Bull.*, 1949, **46**, 198–223.

PEARCE, S. C. Experimenting with organisms as blocks. *Biometrika*, 1957, **44**, 141–149.

PILKINGTON, G. W. Scientific method in psychology. III *Brit. J. statist. Psychol.*, 1958, **11**, 129–132.

ROBSON, D. S. A simple method for constructing orthogonal polynomials when the independent variable is unequally spaced. *Biometrics*, 1959, **15**, 187–191.

ROSENZWEIG, S. An experimental study of "repression" with special reference to need-persistive and ego-defensive reactions to frustration. *J. exp. Psychol.*, 1943, **32**, 64–74.

RYAN, T. A. Multiple comparisons in psychological research. *Psychol Bull*, 1959, **56**, 26–47.

SATTERTHWAITE, F. E. An approximate distribution of estimates of variance components. *Biometrics*, 1946, **2**, 110–114.

SCHEFFE', H. A method for judging all contrasts in the analysis of variance. *Biometrika*, 1953, **40**, 87–104.

SCHROEDER, ELINOR M. *On measurement of motor skills.* New York: King's Crown Press, 1945

SCHULTZ, E. F., JR. Rules of thumb for determining expectations of mean squares. *Biometrics*, 1955, **11**, 123–135.

SIEGEL, S. Nonparametric statistics. *Amer. Statist*, 1957, **11**, 13–19.

SLEIGHT, R. B. The effect of instrument dial shape on legibility. *J. appl Psychol.*, 1948, **32**, 170–188.

SNEDECOR, G. W. *Statistical methods.* (5th ed.) Ames: Iowa State College Press, 1956.

STEVENS, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.) *Handbook of experimental psychology.* New York: Wiley, 1951, 1–49.

TAYLOR, J. G. Scientific method in psychology. IV. *Brit. J. statist Psychol.*, 1958, **11**, 133–136.

TRITES, D. K. Graphic determination of significance of 2×2 contingency tables. *Psychol. Bull*, 1957, **54**, 140–144.

TUKEY, J. W. One degree of freedom for non-additivity. *Biometrics*, 1949, **5**, 232–242.

——— Queries. *Biometrics*, 1955, **11**, 111–113.

UNDERWOOD, B. J. *Psychological research.* New York: Appleton-Century-Crofts, 1957.

VILLARS, D. S. *Statistical design and analysis of experiments for development research.* Dubuque, Iowa: Brown, 1951.

WALKER, HELEN M., and J. LEV. *Statistical inference.* New York: Holt, 1953.

WILK, M. G., and O. KEMPTHORNE. Non-additivities in a Latin square design. *J. Amer. statist. Ass.*, 1957, **52**, 218–236.

WILKS, S S. Weighting systems for linear functions of correlated variables when there is no dependent variable. *Psychometrika*, 1938, **3**, 23–40.

WILLIAMS, E. J. Experimental designs balanced for the estimation of residual effects of treatments. *Austr. J. Sci. Res.*, 1949, **2**, 149–168.

WILLIS, T. R. Scientific method in psychology. I. *Brit. J. statist. Psychol.*, 1958, **11**, 97–104.

WISHART, J. Statistics in agricultural research. *J. R. statist. Soc. Suppl.*, 1934, **1**, 26–51.

———, and T. METAKIDES. Orthogonal polynomial fitting. *Biometrika*, 1953, **40**, 361–369.

WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

WYATT, R. F. Improvability of pitch discrimination. *Psychol. Monogr.*, 1945, No. 267.

LIST OF FORMULAS

The numbers given in the parentheses are used throughout the text to refer to the formula. The page on which the formula appears in the text is given at the left.

Page	Number	Formula
16	(2.1)	$_nP_n = n!$
16	(2.2)	$_nP_r = \dfrac{n!}{(n-r)!}$
17	(2.3)	$_nC_r = \dfrac{_nP_r}{_rP_r} = \dfrac{\dfrac{n!}{(n-r)!}}{r!} = \dfrac{n!}{r!(n-r)!}$
23	(2.4)	$_nP_{r_1, r_2, \cdots, r_k} = \dfrac{n!}{r_1!\, r_2! \cdots r_k!}$
32	(3.1)	$m = \dfrac{\sum X}{N}$
32	(3.2)	$m = \dfrac{\sum FX}{N}$
32	(3.3)	$m = \dfrac{\sum FX}{N} = \dfrac{F_1}{N} = P$
33	(3.4)	$\sigma^2 = \dfrac{\sum (X-m)^2}{N}$
33	(3.5)	$\sigma^2 = \dfrac{\sum F(X-m)^2}{N}$
33	(3.6)	$\sigma = \sqrt{\dfrac{\sum (X-m)^2}{N}}$
33	(3.7)	$\sigma = \sqrt{\dfrac{\sum F(X-m)^2}{N}}$

Page	Number	Formula
33	(3.8)	$\sigma = \sqrt{PQ}$
35	(3.9)	$m = P = \frac{\sum Fp}{N}$
36	(3.10)	$\sigma_p^2 = \left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n}$
37	(3.11)	$m = nP$
37	(3.12)	$\sigma_f^2 = \frac{N-n}{N-1} nPQ$
39	(3.13)	$\sigma_p^2 = \frac{\sigma^2}{n}$
39	(3.14)	$\sigma_p = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{PQ}{n}}$
40	(3.15)	$\sigma_f^2 = n^2\sigma_p^2 = n^2\frac{PQ}{n} = nPQ$
40	(3.16)	$\sigma_f = \sqrt{nPQ}$
44	(4.1)	$x = X - m$
44	(4.2)	$z = \frac{X-m}{\sigma}$
44	(4.3)	$y = \frac{1}{\sqrt{2\pi}} e^{-(\frac{1}{2})z^2}$
46	(4.4)	$z = \frac{f-m}{\sigma_f}$
49	(4.5)	$z = \frac{\frac{f}{n} - \frac{nP}{n}}{\sqrt{\frac{nPQ}{n^2}}} = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{p-P}{\sigma_p}$
51	(4 6)	$m = NP = n$
51	(4.7)	$\sigma_f^2 = \frac{N}{N-1} NPQ$

Page	Number	Formula
53	(4.8)	$\sigma_{p_1-p_2} = \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$
53	(4.9)	$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sigma_{p_1-p_2}}$
59	(4.10)	$m = (a + d)P$
59	(4.11)	$\sigma_f = \sqrt{(a + d)PQ}$
59	(4.12)	$z = \frac{d - a}{\sqrt{a + d}}$
59	(4.13)	$z = \frac{\lvert d - a\rvert - 1}{\sqrt{d + a}}$
63	(5.1)	$\chi^2 = \sum_1^c \frac{(f_i - F_i)^2}{F_i}$
64	(5.2)	$F = nP$
66	(5.3)	$\chi^2 = \sum_1^r \sum_1^c \frac{(f_i - F_i)^2}{F_i}$
68	(5.4)	$\chi^2 = \frac{n^2}{(\sum f_1)(\sum f_2)}\left[\sum \frac{f_1^2}{n_i} - \frac{(\sum f_1)^2}{n}\right]$
70	(5.5)	$\chi^2 = \frac{n\left(\lvert bc - ad\rvert - \frac{n}{2}\right)^2}{(a + b)(c + d)(a + c)(b + d)}$
73	(5.6)	$z = \sqrt{2\chi^2} - \sqrt{(2)(\text{d.f.}) - 1}$
77	(6.1)	$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$
78	(6.2)	$t = \frac{r\sqrt{n - 2}}{\sqrt{1 - r^2}}$
79	(6.3)	$z' = \frac{1}{2}[\log_e (1 + r) - \log_e (1 - r)]$
81	(6.4)	$\sigma_z{}' = \frac{1}{\sqrt{n - 3}}$

Page	Number	Formula
81	(6.5)	$z = \dfrac{z' - \bar{z}'}{\sigma_{z'}}$
82	(6.6)	$z' \pm 1.96\sigma_{z'}$
82	(6.7)	$z' \pm 2.58\sigma_{z'}$
82	(6.8)	$\sigma_{z_1' - z_2'} = \sqrt{\dfrac{1}{n_1 - 3} + \dfrac{1}{n_2 - 3}}$
83	(6.9)	$z = \dfrac{(z_1' - z_2') - (\bar{z}_1' - \bar{z}_2')}{\sigma_{z_1' - z_2'}}$
83	(6.10)	$\chi^2 = \sum(n_i - 3)(z_i')^2 - \dfrac{[\sum(n_i - 3)(z_i')]^2}{\sum(n_i - 3)}$
85	(6.11)	$t = (r_1 - r_2)\sqrt{\dfrac{(n-3)(1+r_{12})}{2(1 - r_1^2 - r_2^2 - r_{12}^2 + 2r_1 r_2 r_{12})}}$
86	(7.1)	$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$
87	(7.2)	$\bar{X} = \dfrac{\sum X}{n}$
88	(7.3)	$s^2 = \dfrac{\sum(X - \bar{X})^2}{n - 1}$
88	(7.4)	$s = \sqrt{\dfrac{\sum(X - \bar{X})^2}{n - 1}}$
88	(7.5)	$s_{\bar{x}} = \dfrac{s}{\sqrt{n}}$
88	(7.6)	$t = \dfrac{\bar{X} - m}{s_{\bar{x}}}$
89	(7.7)	$-t \leqq \dfrac{\bar{X} - m}{s_{\bar{x}}} \leqq t$
92	(7.8)	$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2}$
92	(7.9)	$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

Page	Number	Formula
92	(7.10)	$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}$
92	(7.11)	$s^2 = \frac{(n_1-1){s_1}^2 + (n_2-1){s_2}^2 + \cdots + (n_k-1){s_k}^2}{(n_1-1)+(n_2-1)+\cdots+(n_k-1)}$
92	(7.12)	$\sum x^2 = \sum (X - \bar{X})^2$
92	(7.13)	$s^2 = \frac{\sum {x_1}^2 + \sum {x_2}^2 + \cdots \sum {x_k}^2}{\sum n - k}$
92	(7.14)	$s^2 = \frac{\sum {x_1}^2 + \sum {x_2}^2}{n_1 + n_2 - 2}$
93	(7.15)	$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\left(\frac{\sum {x_1}^2 + \sum {x_2}^2}{n_1+n_2-2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$
93	(7.16)	$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{2s^2}{n}} = s\sqrt{\frac{2}{n}}$
93	(7.17)	$\sum x^2 = \sum X^2 - \frac{(\sum X)^2}{n}$
94	(7.18)	$t = \frac{(\bar{X}_1 - \bar{X}_2) - (m_1 - m_2)}{s_{\bar{x}_1-\bar{x}}}$
94	(7.19)	$t = \frac{\bar{X}_1 - \bar{X}_1}{s_{\bar{x}_1-\bar{x}_2}}$
99	(7.20)	$n = 18\frac{\sigma^2}{(m_1-m_2)^2}$
99	(7.21)	$n = 8\frac{\sigma^2}{(m_1-m_2)^2}$
105	(8.1)	$F = \frac{{s_1}^2}{{s_2}^2}$ or $F = \frac{{s_2}^2}{{s_1}^2}$
107	(8.2)	$t_{.05} = \frac{t_1\frac{{s_1}^2}{n_1} + t_2\frac{{s_2}^2}{n_2}}{\frac{{s_1}^2}{n_1}+\frac{{s_2}^2}{n_2}}$

Page	Number	Formula
107	(8.3)	$s_{\bar{x}_1-\bar{x}_2}^2 = \frac{\sum x_1^2 + \sum x_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$
108	(8.4)	$s_{\bar{x}_1-\bar{x}_2}^2 = \frac{\frac{\sum x_1^2}{n_1 - 1}}{n_1} + \frac{\frac{\sum x_2^2}{n_2 - 1}}{n_2}$
119	(9.1)	$\sum x_t^2 = \sum X^2 - \frac{(\sum X)^2}{n}$
119	(9.2)	$\sum x_b^2 = \frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \cdots + \frac{(\sum X_k)^2}{n_k} - \frac{(\sum X)^2}{n}$
119	(9.3)	$\sum x_w^2 = \sum x_t^2 - \sum x_b^2$
121	(9.4)	$F = \frac{\text{Mean square between groups}}{\text{Mean square within groups}}$
123	(9.5)	$\sum_1^k \sum_1^n x^2 = \sum_1^k \sum_1^n x_k^2 + n \sum_1^k a_{k\cdot}^2$
124	(9.6)	$\bar{\bar{X}}_{\cdot\cdot} = \frac{\sum X_{1\cdot} + \sum X_{2\cdot} + \cdots + \sum X_{k\cdot}}{kn}$
124	(9.7)	$s_{\bar{x}}^2 = \frac{\sum_1^k a_{k\cdot}^2}{k - 1}$
124	(9.8)	$ns_{\bar{x}}^2 = s^2$
124	(9.9)	$ns_{\bar{x}}^2 = \frac{n \sum_1^k a_{k\cdot}^2}{k - 1}$
142	(10.1)	$s_{d_i} = \sqrt{s^2\left(\frac{a_{1i}^2}{n_1} + \frac{a_{2i}^2}{n_2} + \cdots + \frac{a_{ki}^2}{n_k}\right)}$
142	(10.2)	$s_{d_i} = \sqrt{\frac{s^2}{n} \sum a_{\cdot i}^2}$
142	(10.3)	$t = \frac{d_i}{s_{d_i}}$

Page	Number	Formula
143	(10.4)	$d_i = a_{1i}\bar{X}_{1\cdot} + a_{2i}\bar{X}_{2\cdot} + \cdots + a_{ki}\bar{X}_{k\cdot}$
144	(10.5)	$a_{1i}a_{1j} + a_{2i}a_{2j} + \cdots + a_{ki}a_{kj} = 0$
144	(10.6)	$d_1 = \frac{1}{n}(a_{11}\sum X_{1\cdot} + a_{21}\sum X_{2\cdot} + \cdots + a_{k1}\sum X_{k\cdot})$
144	(10.7)	$D_1 = a_{11}\sum X_{1\cdot} + a_{21}\sum X_{2\cdot} + \cdots + a_{k1}\sum X_{k\cdot}$
144	(10.8)	$A_1 = \frac{{D_1}^2}{n\sum {a_{\cdot 1}}^2}$
145	(10.9)	$F = \frac{A}{s^2}$
151	(10.10)	$F = \frac{\text{Mean square for deviations from linearity}}{\text{Error mean square}}$
156	(10.11)	$F = \frac{A_i}{s^2}$
156	(10.12)	$F' = (k-1)F$
161	(11.1)	Residual = Total − treatments − blocks
162	(11.2)	$F = \frac{\text{Treatment mean square}}{\text{Residual mean square}}$
167	(11.3)	$\text{Nonadditivity} = \frac{\left[\sum_1^{kn}(X_{kn})(\bar{X}_{k\cdot}-\bar{X}_{\cdot\cdot})(\bar{X}_{\cdot n}-\bar{X}_{\cdot\cdot})\right]^2}{\sum_1^k(\bar{X}_{k\cdot}-\bar{X}_{\cdot\cdot})^2\sum_1^n(\bar{X}_{\cdot n}-\bar{X}_{\cdot\cdot})^2}$
167	(11.4)	$F = \frac{\text{Nonadditivity}}{\text{Remainder}}$
168	(11.5)	Average of the sum of products $\pm 2\sqrt{\begin{pmatrix}\text{Sum of squares of}\\ \text{deviations of}\\ \text{treatment means}\end{pmatrix}\begin{pmatrix}\text{Mean square}\\ \text{for}\\ \text{remainder}\end{pmatrix}}$
171	(11.6)	$\sum(D-\bar{D})^2 = \sum D^2 - \frac{(\sum D)^2}{n}$

Page	Number	Formula
283	(16.2)	$\sum xy_b = \frac{\left(\sum_1^n X_{1n}\right)\left(\sum_1^n Y_{1n}\right)}{n_1} + \frac{\left(\sum_1^n X_{2n}\right)\left(\sum_1^n Y_{2n}\right)}{n_2} + \cdots + \frac{\left(\sum_1^n X_{kn}\right)\left(\sum_1^n Y_{kn}\right)}{n_k} - \frac{\left(\sum_1^{kn} X_{kn}\right)\left(\sum_1^{kn} Y_{kn}\right)}{kn}$
283	(16.3)	$\sum xy_w = \sum_1^k \left[\sum_1^n X_{kn}Y_{kn} - \frac{\left(\sum_1^n X_{kn}\right)\left(\sum_1^n Y_{kn}\right)}{n_k}\right]$
283	(16.4)	$\sum xy_w = \sum xy_t - \sum xy_b$
288	(16.5)	$b_k = \frac{\sum_1^n xy_k}{\sum_1^n x_k^2}$
288	(16.6)	$\tilde{y}_k = b_k x_k$
289	(16.7)	$\sum_1^n (y_k - b_k x_k)^2 = \sum_1^n y_k^2 - \frac{\left(\sum_1^n xy_k\right)^2}{\sum_1^n x_k^2}$
289	(16.8)	$S_1 = \sum y_w^2 - \sum_1^k \frac{\left(\sum_1^n xy_k\right)^2}{\sum_1^n x_k^2}$
290	(16.9)	$b_w = \frac{\sum xy_w}{\sum x_w^2}$
290	(16.10)	$\tilde{y}_k = b_w x_k$
290	(16.11)	$S_2 = \sum y_w^2 - \frac{(\sum xy_w)^2}{\sum x_w^2}$
291	(16.12)	$S_3 = S_2 - S_1$

Page	Number	Formula
292	(16.13)	$F = \dfrac{\dfrac{S_3}{k-1}}{\dfrac{S_1}{k(n-2)}}$
292	(16.14)	$b_t = \dfrac{\sum xy_t}{\sum {x_t}^2}$
292	(16.15)	$\tilde{y} = b_t x$
293	(16.16)	$S_4 = \sum {y_t}^2 - \dfrac{(\sum xy_t)^2}{\sum {x_t}^2}$
293	(16.17)	$S_5 = S_4 - S_2$
293	(16.18)	$S_5 = \sum {y_b}^2 - \left[\dfrac{(\sum xy_t)^2}{\sum {x_t}^2} - \dfrac{(\sum xy_w)^2}{\sum {x_w}^2}\right]$
294	(16.19)	$F = \dfrac{\dfrac{S_5}{k-1}}{\dfrac{S_2}{k(n-2)}}$
295	(16.20)	$D_{kn} = Y_{kn} - X_{kn}$
299	(16.21)	$I = \dfrac{X_1}{s_1} + \dfrac{X_2}{s_2} + \cdots + \dfrac{X_k}{s_k}$

APPENDIX

Table I. Table of Random Numbers*

Row	Column Number							
	00000 01234	00000 56789	11111 01234	11111 56789	22222 01234	22222 56789	33333 01234	33333 56789
	1st Thousand							
00	23157	54859	01837	25993	76249	70886	95230	36744
01	05545	55043	10537	43508	90611	83744	10962	21343
02	14871	60350	32404	36223	50051	00322	11543	80834
03	38976	74951	94051	75853	78805	90194	32428	71695
04	97312	61718	99755	30870	94251	25841	54882	10513
05	11742	69381	44339	30872	32797	33118	22647	06850
06	43361	28859	11016	45623	93009	00499	43640	74036
07	93806	20478	38268	04491	55751	18932	58475	52571
08	49540	13181	08429	84187	69538	29661	77738	09527
09	36768	72633	37948	21569	41959	68670	45274	83880
10	07092	52392	24627	12067	06558	45344	67338	45320
11	43310	01081	44863	80307	52555	16148	89742	94647
12	61570	06360	06173	63775	63148	95123	35017	46993
13	31352	83799	10779	18941	31579	76448	62584	86919
14	57048	86526	27795	93692	90529	56546	35065	32254
15	09243	44200	68721	07137	30729	75756	09298	27650
16	97957	35018	40894	88329	52230	82521	22532	61587
17	93732	59570	43781	98885	56671	66826	95996	44569
18	72621	11225	00922	68264	35666	59434	71687	58167
19	61020	74418	45371	20794	95917	37866	99536	19378
20	97839	85474	33055	91718	45473	54144	22034	23000
21	89160	97192	22232	90637	35055	45489	88438	16361
22	25966	88220	62871	79265	02823	52862	84919	54883
23	81443	31719	05049	54806	74690	07567	65017	16543
24	11322	54931	42362	34386	08624	97687	46245	23245

* Table I is reproduced from M. G. Kendall and B. B. Smith. Randomness and random sampling numbers. *J. R. statist. Soc.*, **101** (1938), 147–166, by permission of the Royal Statistics Society.

Table I. Table of Random Numbers*—Continued

Row	Column Number							
	00000 01234	00000 56789	11111 01234	11111 56789	22222 01234	22222 56789	33333 01234	33333 56789
	2nd Thousand							
00	64755	83885	84122	25920	17696	15655	95045	95947
01	10302	52289	77436	34430	38112	49067	07348	23328
02	71017	98495	51308	50374	66591	02887	53765	69149
03	60012	55605	88410	34879	79655	90169	78800	03666
04	37330	94656	49161	42802	48274	54755	44553	65090
05	47869	87001	31591	12273	60626	12822	34691	61212
06	38040	42737	64167	89578	39323	49324	88434	38706
07	73508	30908	83054	80078	86669	30295	56460	45336
08	32623	46474	84061	04324	20628	37319	32356	43969
09	97591	99549	36630	35106	62069	92975	95320	57734
10	74012	31955	59790	96982	66224	24015	96749	07589
11	56754	26457	13351	05014	90966	33674	69096	33488
12	49800	49908	54831	21998	08528	26372	92923	65026
13	43584	89647	24878	56670	00221	50193	99591	62377
14	16653	79664	60325	71301	35742	83636	73058	87229
15	48502	69055	65322	58748	31446	80237	31252	96367
16	96765	54692	36316	86230	48296	38352	23816	64094
17	38923	61550	80357	81784	23444	12463	33992	28128
18	77958	81694	25225	05587	51073	01070	60218	61961
19	17928	28065	25586	08771	02641	85064	65796	48170
20	94036	85978	02318	04499	41054	10531	87431	21596
21	47460	60479	56230	48417	14372	85167	27558	00368
22	47856	56088	51992	82439	40644	17170	13463	18288
23	57616	34653	92298	62018	10375	76515	62986	90756
24	08300	92704	66752	66610	57188	79107	54222	22013

* Table I is reproduced from M G. Kendall and B. B. Smith. Randomness and random sampling numbers. *J. R. statist. Soc.*, **101** (1938), 147–166, by permission of the Royal Statistical Society.

Table I. Table of Random Numbers*—Continued

Row	Column Number							
	00000 01234	00000 56789	11111 01234	11111 56789	22222 01234	22222 56789	33333 01234	33333 56789
	3rd Thousand							
00	89221	02362	65787	74733	51272	30213	92441	39651
01	04005	99818	63918	29032	94012	42363	01261	10650
02	98546	38066	50856	75045	40645	22841	53254	44125
03	41719	84401	59226	01314	54581	40398	49988	65579
04	28733	72489	00785	25843	24613	49797	85567	84471
05	65213	83927	77762	03086	80742	24395	68476	83792
06	65553	12678	90906	90466	43670	26217	69900	31205
07	05668	69080	73029	85746	58332	78231	45986	92998
08	39302	99718	49757	79519	27387	76373	47262	91612
09	64592	32254	45879	29431	38320	05981	18067	87137
10	07513	48792	47314	83660	68907	05336	82579	91582
11	86593	68501	56638	99800	82839	35148	56541	07232
12	83735	22599	97977	81248	36838	99560	32410	67614
13	08595	21826	54655	08204	87990	17033	56258	05384
14	41273	27149	44293	69458	16828	63962	15864	35431
15	00473	75908	56238	12242	72631	76314	47252	06347
16	86131	53789	81383	07868	89132	96182	07009	86432
17	33849	78359	08402	03586	03176	88663	08018	22546
18	61870	41657	07468	08612	98083	97349	20775	45091
19	43898	65923	25078	86129	78491	97653	91500	80786
20	29939	39123	04548	45985	60952	06641	28726	46473
21	38505	85555	14388	55077	18657	94887	67831	70819
22	31824	38431	67125	25511	72044	11562	53279	82268
23	91430	03767	13561	15597	06750	92552	02391	38753
24	38635	68976	25498	97526	96458	03805	04116	63514

* Table I is reproduced from M. G Kendall and B. B. Smith. Randomness and random sampling numbers. *J. R. statist. Soc.*, **101** (1938), 147–166, by permission of the Royal Statistical Society.

Table I. Table of Random Numbers*—Continued

Row	Column Number							
	00000 01234	00000 56789	11111 01234	11111 56789	22222 01234	22222 56789	33333 01234	33333 56789
	4th Thousand							
00	02490	54122	27944	39364	94239	72074	11679	54082
01	11967	36469	60627	83701	09253	30208	01385	37482
02	48256	83465	49699	24079	05403	35154	39613	03136
03	27246	73080	21481	23536	04881	89977	49484	93071
04	32532	77265	72430	70722	86529	18457	92657	10011
05	66757	98955	92375	93431	43204	55825	45443	69265
06	11266	34545	76505	97746	34668	26999	26742	97516
07	17872	39142	45561	80146	93137	48924	64257	59284
08	62561	30365	03408	14754	51798	08133	61010	97730
09	62796	30779	35497	70501	30105	08133	00997	91970
10	75510	21771	04339	33660	42757	62223	87565	48468
11	87439	01691	63517	26590	44437	07217	98706	39032
12	97742	02621	10748	78803	38337	65226	92149	59051
13	98811	06001	21571	02875	21828	83912	85188	61624
14	51264	01852	64607	92553	29004	26695	78583	62998
15	40239	93376	10419	68610	49120	02941	80035	99317
16	26936	59186	51667	27645	46329	44681	94190	66647
17	88502	11716	98299	40974	42394	62200	69094	81646
18	63499	38093	25593	61995	79867	80569	01023	38374
19	36379	81206	03317	78710	73828	31083	60509	44091
20	93801	22322	47479	57017	59334	30647	43061	26660
21	29856	87120	56311	50053	25365	81265	22414	02431
22	97720	87931	88265	13050	71017	15177	06957	92919
23	85237	09105	74601	46377	59938	15647	34177	92753
24	75746	75268	31727	95773	72364	87324	36879	06802

* Table I is reproduced from M. G Kendall and B. B. Smith. Randomness and random sampling numbers. *J. R. statist. Soc.*, **101** (1938), 147-166, by permission of the Royal Statistical Society.

Table I. Table of Random Numbers*—Concluded

Row	Column Number							
	00000 01234	00000 56789	11111 01234	11111 56789	22222 01234	22222 56789	33333 01234	33333 56789
	5th Thousand							
00	29935	06971	63175	52579	10478	89379	61428	21363
01	15114	07126	51890	77787	75510	13103	42942	48111
02	03870	43225	10589	87629	22039	94124	38127	65022
03	79390	39188	40756	45269	65959	20640	14284	22960
04	30035	06915	79196	54428	64819	52314	48721	81594
05	29039	99861	28759	79802	68531	39198	38137	24373
06	78196	08108	24107	49777	09599	43569	84820	94956
07	15847	85493	91442	91351	80130	73752	21539	10986
08	36614	62248	49194	97209	92587	92053	41021	80064
09	40549	54884	91465	43862	35541	44466	88894	74180
10	40878	08997	14286	09982	90308	78007	51587	16658
11	10229	49282	41173	31468	59455	18756	08908	06660
12	15918	76787	30624	25928	44124	25088	31137	71614
13	13403	18796	49909	94404	64979	41462	18155	98335
14	66523	94596	74908	90271	10009	98648	17640	68909
15	91665	36469	68343	17870	25975	04662	21272	50620
16	67415	87515	08207	73729	73201	57593	96917	69699
17	76527	96996	23724	33448	63392	32394	60887	90617
18	19815	47789	74348	17147	10954	34355	81194	54407
19	25592	53587	76384	72575	84347	68918	05739	57222
20	55902	45539	63646	31609	95999	82887	40666	66692
21	02470	58376	79794	22482	42423	96162	47491	17264
22	18630	53263	13319	97619	35859	12350	14632	87659
23	89673	38230	16063	92007	59503	38402	76450	33333
24	62986	67364	06595	17427	84623	14565	82860	57300

* Table I is reproduced from M. G. Kendall and B. B. Smith. Randomness and random sampling numbers. *J. R. statist. Soc.*, **101** (1938), 147–166, by permission of the Royal Statistical Society.

Table II. Table of Squares, Square Roots, and Reciprocals of Numbers from 1 to 1,000*

N	N^2	$\sqrt{N}$	$1/N$	N	N^2	$\sqrt{N}$	$1/N$
1	1	1.0000	1.000000	41	1681	6.4031	.024390
2	4	1.4142	.500000	42	1764	6.4807	.023810
3	9	1.7321	.333333	43	1849	6.5574	.023256
4	16	2.0000	.250000	44	1936	6.6332	.022727
5	25	2.2361	.200000	45	2025	6.7082	.022222
6	36	2.4495	.166667	46	2116	6.7823	.021739
7	49	2.6458	.142857	47	2209	6.8557	.021277
8	64	2.8284	.125000	48	2304	6.9282	.020833
9	81	3.0000	.111111	49	2401	7.0000	.020408
10	100	3.1623	.100000	50	2500	7.0711	.020000
11	121	3.3166	.090909	51	2601	7.1414	.019608
12	144	3.4641	.083333	52	2704	7.2111	.019231
13	169	3.6056	.076923	53	2809	7.2801	.018868
14	196	3.7417	.071429	54	2916	7.3485	.018519
15	225	3.8730	.066667	55	3025	7.4162	.018182
16	256	4.0000	.062500	56	3136	7.4833	.017857
17	289	4.1231	.058824	57	3249	7.5498	.017544
18	324	4.2426	.055556	58	3364	7.6158	.017241
19	361	4.3589	.052632	59	3481	7.6811	.016949
20	400	4.4721	.050000	60	3600	7.7460	.016667
21	441	4.5826	.047619	61	3721	7.8102	.016393
22	484	4.6904	.045455	62	3844	7.8740	.016129
23	529	4.7958	.043478	63	3969	7.9373	.015873
24	576	4.8990	.041667	64	4096	8.0000	.015625
25	625	5.0000	.040000	65	4225	8.0623	.015385
26	676	5.0990	.038462	66	4356	8.1240	.015152
27	729	5.1962	.037037	67	4489	8.1854	.014925
28	784	5.2915	.035714	68	4624	8.2462	.014706
29	841	5.3852	.034483	69	4761	8.3066	.014493
30	900	5.4772	.033333	70	4900	8.3666	.014286
31	961	5.5678	.032258	71	5041	8.4261	.014085
32	1024	5.6569	.031250	72	5184	8.4853	.013889
33	1089	5.7446	.030303	73	5329	8.5440	.013699
34	1156	5.8310	.029412	74	5476	8.6023	.013514
35	1225	5.9161	.028571	75	5625	8.6603	.013333
36	1296	6.0000	.027778	76	5776	8.7178	.013158
37	1369	6.0828	.027027	77	5929	8.7750	.012987
38	1444	6.1644	.026316	78	6084	8.8318	.012821
39	1521	6.2450	.025641	79	6241	8.8882	.012658
40	1600	6.3246	.025000	80	6400	8.9443	.012500

* Portions of Table II have been reproduced from J. W. Dunlap and A. K. Kurtz, *Handbook of Statistical Nomographs, Tables, and Formulas*, World Book Company, New York (1932), by permission of the authors and publishers.

Table II. Table of Squares, Square Roots, and Reciprocals of Numbers from 1 to 1,000*—Continued

N	N^2	$\sqrt{N}$	$1/N$	N	N^2	$\sqrt{N}$	$1/N$
81	6561	9.0000	.012346	121	14641	11.0000	.00826446
82	6724	9.0554	.012195	122	14884	11.0454	00819672
83	6889	9.1104	.012048	123	15129	11.0905	.00813008
84	7056	9.1652	.011905	124	15376	11.1355	.00806452
85	7225	9.2195	.011765	125	15625	11.1803	.00800000
86	7396	9.2736	.011628	126	15876	11.2250	.00793651
87	7569	9.3274	.011494	127	16129	11 2694	.00787402
88	7744	9.3808	.011364	128	16384	11.3137	.00781250
89	7921	9 4340	.011236	129	16641	11.3578	.00775194
90	8100	9.4868	.011111	130	16900	11 4018	.00769231
91	8281	9.5394	.010989	131	17161	11.4455	.00763359
92	8464	9 5917	.010870	132	17424	11.4891	.00757576
93	8649	9.6437	.010753	133	17689	11.5326	.00751880
94	8836	9.6954	.010638	134	17956	11.5758	.00746269
95	9025	9.7468	.010526	135	18225	11.6190	.00740741
96	9216	9.7980	.010417	136	18496	11.6619	.00735294
97	9409	9.8489	.010309	137	18769	11.7047	.00729927
98	9604	9.8995	.010204	138	19044	11.7473	.00724638
99	9801	9.9499	.010101	139	19321	11.7898	.00719424
100	10000	10.0000	.010000	140	19600	11 8322	.00714286
101	10201	10.0499	.00990099	141	19881	11.8743	.00709220
102	10404	10 0995	.00980392	142	20164	11.9164	.00704225
103	10609	10.1489	.00970874	143	20449	11.9583	.00699301
104	10816	10.1980	.00961538	144	20736	12.0000	.00694444
105	11025	10.2470	.00952381	145	21025	12 0416	.00689655
106	11236	10.2956	.00943396	146	21316	12.0830	.00684932
107	11449	10.3441	.00934579	147	21609	12 1244	.00680272
108	11664	10.3923	.00925926	148	21904	12.1655	.00675676
109	11881	10.4403	00917431	149	22201	12.2066	.00671141
110	12100	10.4881	.00909091	150	22500	12.2474	.00666667
111	12321	10.5357	.00900901	151	22801	12.2882	.00662252
112	12544	10 5830	.00892857	152	23104	12.3288	.00657895
113	12769	10 6301	.00884956	153	23409	12.3693	.00653595
114	12996	10.6771	.00877193	154	23716	12.4097	.00649351
115	13225	10.7238	.00869565	155	24025	12.4499	.00645161
116	13456	10 7703	.00862069	156	24336	12.4900	.00641026
117	13689	10.8167	.00854701	157	24649	12.5300	.00636943
118	13924	10.8628	00847458	158	24964	12.5698	.00632911
119	14161	10.9087	.00840336	159	25281	12.6095	.00628931
120	14400	10 9545	00833333	160	25600	12 6491	.00625000

* Portions of Table II have been reproduced from J. W. Dunlap and A K. Kurtz *Handbook of Statistical Nomographs, Tables, and Formulas*, World Book Company, New York (1932), by permission of the authors and publishers.

Table II. Table of Squares, Square Roots, and Reciprocals of Numbers from 1 to 1,000*—Continued

N	N^2	$\sqrt{N}$	$1/N$	N	N^2	$\sqrt{N}$	$1/N$
161	25921	12 6886	.00621118	201	40401	14.1774	.00497512
162	26244	12.7279	.00617284	202	40804	14.2127	.00495050
163	26569	12.7671	.00613497	203	41209	14.2478	.00492611
164	26896	12.8062	.00609756	204	41616	14.2829	.00490196
165	27225	12.8452	.00606061	205	42025	14 3178	.00487805
166	27556	12.8841	.00602410	206	42436	14.3527	.00485437
167	27889	12.9228	.00598802	207	42849	14.3875	.00483092
168	28224	12.9615	.00595238	208	43264	14.4222	.00480769
169	28561	13.0000	.00591716	209	43681	14.4568	.00478469
170	28900	13.0384	.00588235	210	44100	14.4914	.00476190
171	29241	13.0767	.00584795	211	44521	14.5258	.00473934
172	29584	13 1149	.00581395	212	44944	14.5602	.00471698
173	29929	13.1529	.00578035	213	45369	14 5945	.00469484
174	30276	13.1909	.00574713	214	45796	14.6287	.00467290
175	30625	13.2288	.00571429	215	46225	14 6629	.00465116
176	30976	13.2665	.00568182	216	46656	14.6969	.00462963
177	31329	13.3041	.00564972	217	47089	14.7309	.00460829
178	31684	13.3417	.00561798	218	47524	14.7648	.00458716
179	32041	13.3791	.00558659	219	47961	14.7986	.00456621
180	32400	13.4164	.00555556	220	48400	14.8324	.00454545
181	32761	13.4536	.00552486	221	48841	14.8661	.00452489
182	33124	13.4907	.00549451	222	49284	14.8997	.00450450
183	33489	13.5277	.00546448	223	49729	14.9332	.00448430
184	33856	13.5647	.00543478	224	50176	14.9666	.00446429
185	34225	13.6015	.00540541	225	50625	15.0000	.00444444
186	34596	13.6382	.00537634	226	51076	15.0333	.00442478
187	34969	13.6748	.00534759	227	51529	15.0665	.00440529
188	35344	13.7113	.00531915	228	51984	15.0997	.00438596
189	35721	13.7477	.00529101	229	52441	15.1327	.00436681
190	36100	13.7840	.00526316	230	52900	15.1658	.00434783
191	36481	13.8203	.00523560	231	53361	15.1987	.00432900
192	36864	13.8564	.00520833	232	53824	15.2315	.00431034
193	37249	13 8924	.00518135	233	54289	15.2643	.00429185
194	37636	13.9284	.00515464	234	54756	15.2971	.00427350
195	38025	13.9642	.00512821	235	55225	15.3297	.00425532
196	38416	14.0000	.00510204	236	55696	15.3623	.00423729
197	38809	14.0357	.00507614	237	56169	15.3948	.00421941
198	39204	14 0712	.00505051	238	56644	15.4272	.00420168
199	39601	14.1067	.00502513	239	57121	15.4596	.00418410
200	40000	14 1421	.00500000	240	57600	15.4919	.00416667

* Portions of Table II have been reproduced from J. W Dunlap and A. K. Kurtz. *Handbook of Statistical Nomographs, Tables, and Formulas*, World Book Company, New York (1932), by permission of the authors and publishers.

Table II. Table of Squares, Square Roots, and Reciprocals of Numbers from 1 to 1,000*—Continued

N	N^2	$\sqrt{N}$	$1/N$	N	N^2	$\sqrt{N}$	$1/N$
241	58081	15.5242	.00414938	281	78961	16.7631	.00355872
242	58564	15.5563	.00413223	282	79524	16.7929	.00354610
243	59049	15.5885	.00411523	283	80089	16.8226	.00353357
244	59536	15.6205	.00409836	284	80656	16.8523	.00352113
245	60025	15.6525	.00408163	285	81225	16.8819	.00350877
246	60516	15.6844	.00406504	286	81796	16.9115	.00349650
247	61009	15.7162	.00404858	287	82369	16.9411	.00348432
248	61504	15.7480	.00403226	288	82944	16.9706	.00347222
249	62001	15.7797	.00401606	289	83521	17.0000	.00346021
250	62500	15.8114	.00400000	290	84100	17 0294	.00344828
251	63001	15.8430	.00398406	291	84681	17.0587	.00343643
252	63504	15.8745	.00396825	292	85264	17.0880	.00342466
253	64009	15.9060	.00395257	293	85849	17.1172	.00341297
254	64516	15.9374	.00393701	294	86436	17.1464	.00340136
255	65025	15.9687	.00392157	295	87025	17.1756	.00338983
256	65536	16.0000	.00390625	296	87616	17.2047	.00337838
257	66049	16.0312	.00389105	297	88209	17.2337	.00336700
258	66564	16.0624	.00387597	298	88804	17.2627	.00335570
259	67081	16.0935	.00386100	299	89401	17.2916	.00334448
260	67600	16.1245	.00384615	300	90000	17.3205	.00333333
261	68121	16.1555	.00383142	301	90601	17.3494	.00332226
262	68644	16.1864	.00381679	302	91204	17.3781	.00331126
263	69169	16.2173	.00380228	303	91809	17.4069	.00330033
264	69696	16.2481	.00378788	304	92416	17.4356	.00328947
265	70225	16.2788	.00377358	305	93025	17.4642	.00327869
266	70756	16.3095	.00375940	306	93636	17.4929	.00326797
267	71289	16.3401	.00374532	307	94249	17.5214	.00325733
268	71824	16.3707	.00373134	308	94864	17.5499	.00324675
269	72361	16.4012	.00371747	309	95481	17.5784	00323625
270	72900	16.4317	.00370370	310	96100	17.6068	.00322581
271	73441	16.4621	.00369004	311	96721	17.6352	.00321543
272	73984	16.4924	.00367647	312	97344	17.6635	.00320513
273	74529	16.5227	.00366300	313	97969	17.6918	.00319489
274	75076	16.5529	.00364964	314	98596	17.7200	.00318471
275	75625	16.5831	.00363636	315	99225	17 7482	.00317460
276	76176	16.6132	.00362319	316	99856	17.7764	.00316456
277	76729	16.6433	.00361011	317	100489	17.8045	.00315457
278	77284	16.6733	.00359712	318	101124	17.8326	.00314465
279	77841	16.7033	.00358423	319	101761	17.8606	.00313480
280	78400	16.7332	.00357143	320	102400	17.8885	.00312500

* Portions of Table II have been reproduced from J. W. Dunlap and A. K. Kurtz, *Handbook of Statistical Nomographs, Tables, and Formulas*, World Book Company, New York (1932), by permission of the authors and publishers.

Table II. Table of Squares, Square Roots, and Reciprocals of Numbers from 1 to 1,000*—Continued

N	N^2	$\sqrt{N}$	$1/N$	N	N^2	$\sqrt{N}$	$1/N$
321	103041	17.9165	.00311526	361	130321	19.0000	.00277008
322	103684	17.9444	.00310559	362	131044	19.0263	.00276243
323	104329	17.9722	.00309598	363	131769	19.0526	.00275482
324	104976	18.0000	.00308642	364	132496	19.0788	.00274725
325	105625	18.0278	.00307692	365	133225	19.1050	.00273973
326	106276	18.0555	.00306748	366	133956	19.1311	.00273224
327	106929	18.0831	.00305810	367	134689	19.1572	.00272480
328	107584	18.1108	.00304878	368	135424	19.1833	.00271739
329	108241	18.1384	.00303951	369	136161	19.2094	.00271003
330	108900	18.1659	.00303030	370	136900	19.2354	.00270270
331	109561	18.1934	.00302115	371	137641	19.2614	.00269542
332	110224	18.2209	.00301205	372	138384	19.2873	.00268817
333	110889	18.2483	.00300300	373	139129	19.3132	.00268097
334	111556	18.2757	.00299401	374	139876	19.3391	.00267380
335	112225	18.3030	.00298507	375	140625	19.3649	.00266667
336	112896	18.3303	.00297619	376	141376	19.3907	.00265957
337	113569	18.3576	.00296736	377	142129	19.4165	.00265252
338	114244	18.3848	.00295858	378	142884	19.4422	.00264550
339	114921	18.4120	.00294985	379	143641	19.4679	.00263852
340	115600	18.4391	.00294118	380	144400	19.4936	.00263158
341	116281	18.4662	.00293255	381	145161	19.5192	.00262467
342	116964	18.4932	.00292398	382	145924	19.5448	.00261780
343	117649	18.5203	.00291545	383	146689	19.5704	.00261097
344	118336	18.5472	.00290698	384	147456	19.5959	.00260417
345	119025	18.5742	.00289855	385	148225	19.6214	.00259740
346	119716	18.6011	.00289017	386	148996	19.6469	.00259067
347	120409	18.6279	.00288184	387	149769	19.6723	.00258398
348	121104	18.6548	.00287356	388	150544	19.6977	.00257732
349	121801	18.6815	.00286533	389	151321	19.7231	.00257069
350	122500	18.7083	.00285714	390	152100	19.7484	.00256410
351	123201	18.7350	.00284900	391	152881	19.7737	.00255754
352	123904	18.7617	.00284091	392	153664	19.7990	.00255102
353	124609	18.7883	.00283286	393	154449	19.8242	.00254453
354	125316	18.8149	.00282486	394	155236	19.8494	.00253807
355	126025	18.8414	.00281690	395	156025	19.8746	.00253165
356	126736	18.8680	.00280899	396	156816	19.8997	.00252525
357	127449	18.8944	.00280112	397	157609	19.9249	.00251889
358	128164	18.9209	.00279330	398	158404	19.9499	.00251256
359	128881	18.9473	.00278552	399	159201	19.9750	.00250627
360	129600	18.9737	.00277778	400	160000	20.0000	.00250000

* Portions of Table II have been reproduced from J. W. Dunlap and A. K. Kurtz. *Handbook of Statistical Nomographs, Tables, and Formulas*, World Book Company, New York (1932), by permission of the authors and publishers.

Table II. Table of Squares, Square Roots, and Reciprocals of Numbers from 1 to 1,000*—Continued

N	N^2	$\sqrt{N}$	$1/N$	N	N^2	$\sqrt{N}$	$1/N$
401	160801	20.0250	.00249377	441	194481	21.0000	.00226757
402	161604	20.0499	00248756	442	195364	21.0238	.00226244
403	162409	20.0749	.00248139	443	196249	21.0476	.00225734
404	163216	20.0998	.00247525	444	197136	21.0713	.00225225
405	164025	20.1246	.00246914	445	198025	21.0950	.00224719
406	164836	20.1494	.00246305	446	198916	21.1187	.00224215
407	165649	20.1742	.00245700	447	199809	21.1424	.00223714
408	166464	20.1990	.00245098	448	200704	21.1660	.00223214
409	167281	20.2237	.00244499	449	201601	21.1896	.00222717
410	168100	20.2485	.00243902	450	202500	21.2132	.00222222
411	168921	20.2731	.00243309	451	203401	21.2368	.00221729
412	169744	20.2978	.00242718	452	204304	21.2603	.00221239
413	170569	20.3224	.00242131	453	205209	21.2838	.00220751
414	171396	20.3470	.00241546	454	206116	21.3073	.00220264
415	172225	20.3715	.00240964	455	207025	21.3307	.00219780
416	173056	20.3961	.00240385	456	207936	21.3542	.00219298
417	173889	20.4206	.00239808	457	208849	21.3776	.00218818
418	174724	20.4450	.00239234	458	209764	21.4009	.00218341
419	175561	20.4695	.00238663	459	210681	21.4243	.00217865
420	176400	20.4939	.00238095	460	211600	21.4476	.00217391
421	177241	20.5183	.00237530	461	212521	21.4709	.00216920
422	178084	20.5426	.00236967	462	213444	21.4942	.00216450
423	178929	20.5670	.00236407	463	214369	21.5174	.00215983
424	179776	20.5913	.00235849	464	215296	21.5407	.00215517
425	180625	20.6155	.00235294	465	216225	21.5639	.00215054
426	181476	20.6398	.00234742	466	217156	21.5870	.00214592
427	182329	20.6640	.00234192	467	218089	21.6102	.00214133
428	183184	20.6882	.00233645	468	219024	21.6333	.00213675
429	184041	20.7123	.00233100	469	219961	21.6564	.00213220
430	184900	20.7364	.00232558	470	220900	21.6795	.00212766
431	185761	20.7605	.00232019	471	221841	21.7025	.00212314
432	186624	20.7846	.00231481	472	222784	21.7256	.00211864
433	187489	20.8087	.00230947	473	223729	21.7486	.00211416
434	188356	20.8327	.00230415	474	224676	21.7715	.00210970
435	189225	20.8567	.00229885	475	225625	21.7945	.00210526
436	190096	20.8806	.00229358	476	226576	21.8174	.00210084
437	190969	20.9045	.00228833	477	227529	21.8403	.00209644
438	191844	20.9284	.00228311	478	228484	21.8632	.00209205
439	192721	20.9523	.00227790	479	229441	21.8861	.00208768
440	193600	20.9762	.00227273	480	230400	21 9089	.00208333

* Portions of Table II have been reproduced from J. W. Dunlap and A. K. Kurtz, *Handbook of Statistical Nomographs, Tables, and Formulas*, World Book Company, New York (1932), by permission of the authors and publishers.

Table II. Table of Squares, Square Roots, and Reciprocals of Numbers from 1 to 1,000*—Continued

N	N^2	$\sqrt{N}$	$1/N$	N	N^2	$\sqrt{N}$	$1/N$
481	231361	21.9317	.00207900	521	271441	22.8254	.00191939
482	232324	21.9545	.00207469	522	272484	22.8473	.00191571
483	233289	21.9773	.00207039	523	273529	22.8692	.00191205
484	234256	22.0000	.00206612	524	274576	22 8910	.00190840
485	235225	22.0227	.00206186	525	275625	22.9129	.00190476
486	236196	22.0454	.00205761	526	276676	22.9347	.00190114
487	237169	22.0681	.00205339	527	277729	22.9565	.00189753
488	238144	22.0907	.00204918	528	278784	22.9783	.00189394
489	239121	22.1133	.00204499	529	279841	23.0000	.00189036
490	240100	22.1359	.00204082	530	280900	23.0217	.00188679
491	241081	22.1585	.00203666	531	281961	23.0434	.00188324
492	242064	22.1811	.00203252	532	283024	23.0651	.00187970
493	243049	22.2036	.00202840	533	284089	23.0868	.00187617
494	244036	22.2261	.00202429	534	285156	23.1084	.00187266
495	245025	22.2486	.00202020	535	286225	23.1301	.00186916
496	246016	22.2711	.00201613	536	287296	23.1517	.00186567
497	247009	22.2935	.00201207	537	288369	23.1733	.00186220
498	248004	22.3159	.00200803	538	289444	23.1948	.00185874
499	249001	22.3383	.00200401	539	290521	23.2164	.00185529
500	250000	22.3607	.00200000	540	291600	23.2379	.00185185
501	251001	22.3830	.00199601	541	292681	23.2594	.00184843
502	252004	22.4054	.00199203	542	293764	23.2809	.00184502
503	253009	22.4277	.00198807	543	294849	23.3024	.00184162
504	254016	22.4499	.00198413	544	295936	23.3238	.00183824
505	255025	22.4722	.00198020	545	297025	23.3452	.00183486
506	256036	22.4944	.00197628	546	298116	23.3666	.00183150
507	257049	22.5167	.00197239	547	299209	23.3880	.00182815
508	258064	22.5389	.00196850	548	300304	23.4094	.00182482
509	259081	22.5610	.00196464	549	301401	23.4307	.00182149
510	260100	22.5832	.00196078	550	302500	23.4521	.00181818
511	261121	22.6053	.00195695	551	303601	23.4734	.00181488
512	262144	22.6274	.00195312	552	304704	23.4947	.00181159
513	263169	22.6495	.00194932	553	305809	23.5160	.00180832
514	264196	22.6716	.00194553	554	306916	23.5372	.00180505
515	265225	22.6936	.00194175	555	308025	23.5584	.00180180
516	266256	22.7156	.00193798	556	309136	23.5797	.00179856
517	267289	22.7376	.00193424	557	310249	23.6008	.00179533
518	268324	22.7596	.00193050	558	311364	23.6220	.00179211
519	269361	22.7816	.00192678	559	312481	23 6432	.00178891
520	270400	22 8035	.00192308	560	313600	23 6643	.00178571

* Portions of Table II have been reproduced from J. W. Dunlap and A. K. Kurtz. *Handbook of Statistical Nomographs, Tables, and Formulas*, World Book Company, New York (1932), by permission of the authors and publishers.

Table II. Table of Squares, Square Roots, and Reciprocals of Numbers from 1 to 1,000*—Continued

N	N^2	$\sqrt{N}$	$1/N$	N	N^2	$\sqrt{N}$	$1/N$
561	314721	23.6854	.00178253	601	361201	24.5153	.00166389
562	315844	23.7065	.00177936	602	362404	24.5357	.00166113
563	316969	23.7276	.00177620	603	363609	24 5561	.00165837
564	318096	23.7487	.00177305	604	364816	24.5764	.00165563
565	319225	23.7697	.00176991	605	366025	24.5967	.00165289
566	320356	23.7908	.00176678	606	367236	24.6171	.00165017
567	321489	23.8118	.00176367	607	368449	24.6374	.00164745
568	322624	23.8328	.00176056	608	369664	24.6577	.00164474
569	323761	23.8537	.00175747	609	370881	24.6779	.00164204
570	324900	23.8747	.00175439	610	372100	24.6982	.00163934
571	326041	23.8956	.00175131	611	373321	24.7184	.00163666
572	327184	23.9165	.00174825	612	374544	24.7386	.00163399
573	328329	23.9374	.00174520	613	375769	24.7588	.00163132
574	329476	23.9583	.00174216	614	376996	24.7790	.00162866
575	330625	23.9792	.00173913	615	378225	24.7992	.00162602
576	331776	24.0000	.00173611	616	379456	24.8193	.00162338
577	332929	24.0208	.00173310	617	380689	24.8395	.00162075
578	334084	24.0416	.00173010	618	381924	24.8596	.00161812
579	335241	24.0624	.00172712	619	383161	24.8797	.00161551
580	336400	24.0832	.00172414	620	384400	24.8998	.00161290
581	337561	24.1039	.00172117	621	385641	24.9199	.00161031
582	338724	24.1247	.00171821	622	386884	24.9399	.00160772
583	339889	24.1454	.00171527	623	388129	24.9600	.00160514
584	341056	24.1661	.00171233	624	389376	24.9800	.00160256
585	342225	24.1868	.00170940	625	390625	25.0000	.00160000
586	343396	24.2074	.00170648	626	391876	25.0200	.00159744
587	344569	24.2281	.00170358	627	393129	25.0400	.00159490
588	345744	24.2487	.00170068	628	394384	25.0599	.00159236
589	346921	24.2693	.00169779	629	395641	25.0799	.00158983
590	348100	24.2899	.00169492	630	396900	25.0998	.00158730
591	349281	24.3105	.00169205	631	398161	25.1197	.00158479
592	350464	24.3311	.00168919	632	399424	25.1396	.00158228
593	351649	24.3516	.00168634	633	400689	25.1595	.00157978
594	352836	24.3721	.00168350	634	401956	25.1794	.00157729
595	354025	24 3926	.00168067	635	403225	25.1992	.00157480
596	355216	24.4131	.00167785	636	404496	25.2190	.00157233
597	356409	24.4336	.00167504	637	405769	25.2389	.00156986
598	357604	24.4540	.00167224	638	407044	25.2587	.00156740
599	358801	24.4745	.00166945	639	408321	25.2784	.00156495
600	360000	24.4949	.00166667	640	409600	25.2982	.00156250

* Portions of Table II have been reproduced from J. W. Dunlap and A. K. Kurtz. *Handbook of Statistical Nomographs, Tables, and Formulas*, World Book Company, New York (1932), by permission of the authors and publishers.

Table II. Table of Squares, Square Roots, and Reciprocals of Numbers from 1 to 1,000*—Continued

N	N^2	$\sqrt{N}$	$1/N$	N	N^2	$\sqrt{N}$	$1/N$
641	410881	25.3180	.00156006	681	463761	26.0960	.00146843
642	412164	25.3377	.00155763	682	465124	26 1151	.00146628
643	413449	25.3574	.00155521	683	466489	26.1343	.00146413
644	414736	25.3772	.00155280	684	467856	26.1534	.00146199
645	416025	25.3969	.00155039	685	469225	26.1725	.00145985
646	417316	25.4165	.00154799	686	470596	26.1916	.00145773
647	418609	25.4362	.00154560	687	471969	26.2107	.00145560
648	419904	25.4558	.00154321	688	473344	26.2298	.00145349
649	421201	25.4755	.00154083	689	474721	26.2488	.00145138
650	422500	25.4951	.00153846	690	476100	26.2679	.00144928
651	423801	25.5147	.00153610	691	477481	26.2869	.00144718
652	425104	25.5343	.00153374	692	478864	26.3059	.00144509
653	426409	25.5539	.00153139	693	480249	26.3249	.00144300
654	427716	25.5734	.00152905	694	481636	26.3439	.00144092
655	429025	25.5930	.00152672	695	483025	26.3629	.00143885
656	430336	25.6125	.00152439	696	484416	26.3818	.00143678
657	431649	25.6320	.00152207	697	485809	26.4008	.00143472
658	432964	25.6515	.00151976	698	487204	26.4197	.00143266
659	434281	25.6710	.00151745	699	488601	26.4386	.00143062
660	435600	25.6905	.00151515	700	490000	26.4575	.00142857
661	436921	25.7099	.00151286	701	491401	26.4764	.00142653
662	438244	25.7294	.00151057	702	492804	26.4953	.00142450
663	439569	25.7488	.00150830	703	494209	26.5141	.00142248
664	440896	25.7682	.00150602	704	495616	26.5330	.00142045
665	442225	25.7876	.00150376	705	497025	26.5518	.00141844
666	443556	25.8070	.00150150	706	498436	26.5707	.00141643
667	444889	25.8263	.00149925	707	499849	26.5895	.00141443
668	446224	25.8457	.00149701	708	501264	26.6083	.00141243
669	447561	25.8650	.00149477	709	502681	26.6271	.00141044
670	448900	25.8844	.00149254	710	504100	26.6458	.00140845
671	450241	25.9037	.00149031	711	505521	26.6646	.00140647
672	451584	25.9230	.00148810	712	506944	26.6833	.00140449
673	452929	25.9422	.00148588	713	508369	26.7021	.00140252
674	454276	25.9615	.00148368	714	509796	26.7208	.00140056
675	455625	25.9808	.00148148	715	511225	26.7395	.00139860
676	456976	26.0000	.00147929	716	512656	26.7582	.00139665
677	458329	26.0192	.00147710	717	514089	26.7769	.00139470
678	459684	26 0384	.00147493	718	515524	26.7955	.00139276
679	461041	26.0576	.00147275	719	516961	26.8142	.00139082
680	462400	26.0768	.00147059	720	518400	26 8328	.00138889

* Portions of Table II have been reproduced from J. W. Dunlap and A. K. Kurtz. *Handbook of Statistical Nomographs, Tables, and Formulas*, World Book Company, New York (1932), by permission of the authors and publishers.

Table II. Table of Squares, Square Roots, and Reciprocals of Numbers from 1 to 1,000*—Continued

N	N^2	$\sqrt{N}$	$1/N$	N	N^2	$\sqrt{N}$	$1/N$
721	519841	26 8514	.00138696	761	579121	27.5862	.00131406
722	521284	26.8701	.00138504	762	580644	27.6043	.00131234
723	522729	26 8887	.00138313	763	582169	27 6225	.00131062
724	524176	26.9072	.00138122	764	583696	27.6405	.00130890
725	525625	26 9258	.00137931	765	585225	27.6586	.00130719
726	527076	26.9444	.00137741	766	586756	27.6767	.00130548
727	528529	26.9629	.00137552	767	588289	27.6948	.00130378
728	529984	26.9815	.00137363	768	589824	27.7128	.00130208
729	531441	27.0000	.00137174	769	591361	27.7308	.00130039
730	532900	27.0185	.00136986	770	592900	27.7489	.00129870
731	534361	27.0370	.00136799	771	594441	27.7669	.00129702
732	535824	27.0555	.00136612	772	595984	27.7849	.00129534
733	537289	27.0740	.00136426	773	597529	27 8029	.00129366
734	538756	27.0924	.00136240	774	599076	27.8209	.00129199
735	540225	27.1109	.00136054	775	600625	27.8388	.00129032
736	541696	27.1293	.00135870	776	602176	27.8568	.00128866
737	543169	27.1477	.00135685	777	603729	27.8747	.00128700
738	544644	27.1662	.00135501	778	605284	27.8927	.00128535
739	546121	27.1846	.00135318	779	606841	27.9106	.00128370
740	547600	27.2029	.00135135	780	608400	27.9285	.00128205
741	549081	27.2213	.00134953	781	609961	27.9464	.00128041
742	550564	27.2397	.00134771	782	611524	27 9643	.00127877
743	552049	27.2580	.00134590	783	613089	27 9821	.00127714
744	553536	27.2764	.00134409	784	614656	28.0000	.00127551
745	555025	27.2947	.00134228	785	616225	28.0179	.00127389
746	556516	27.3130	.00134048	786	617796	28 0357	.00127226
747	558009	27 3313	.00133869	787	619369	28.0535	.00127065
748	559504	27.3496	.00133690	788	620944	28 0713	.00126904
749	561001	27.3679	.00133511	789	622521	28.0891	00126743
750	562500	27.3861	.00133333	790	624100	28.1069	.00126582
751	564001	27 4044	.00133156	791	625681	28.1247	.00126422
752	565504	27.4226	.00132979	792	627264	28.1425	.00126263
753	567009	27 4408	.00132802	793	628849	28.1603	.00126103
754	568516	27.4591	.00132626	794	630436	28.1780	.00125945
755	570025	27.4773	00132450	795	632025	28.1957	.00125786
756	571536	27 4955	.00132275	796	633616	28.2135	.00125628
757	573049	27.5136	00132100	797	635209	28.2312	.00125471
758	574564	27.5318	.00131926	798	636804	28 2489	.00125313
759	576081	27 5500	.00131752	799	638401	28.2666	.00125156
760	577600	27.5681	.00131579	800	640000	28.2843	.00125000

* Portions of Table II have been reproduced from J. W. Dunlap and A. K. Kurtz. *Handbook of Statistical Nomographs, Tables, and Formulas*, World Book Company, New York (1932), by permission of the authors and publishers.

Table II. Table of Squares, Square Roots, and Reciprocals of Numbers from 1 to 1,000*—Continued

N	N^2	$\sqrt{N}$	$1/N$	N	N^2	$\sqrt{N}$	$1/N$
801	641601	28.3019	.00124844	841	707281	29.0000	.00118906
802	643204	28.3196	.00124688	842	708964	29.0172	.00118765
803	644809	28.3373	.00124533	843	710649	29.0345	.00118624
804	646416	28.3549	.00124378	844	712336	29.0517	.00118483
805	648025	28.3725	.00124224	845	714025	29.0689	.00118343
806	649636	28.3901	.00124069	846	715716	29.0861	.00118203
807	651249	28.4077	.00123916	847	717409	29.1033	00118064
808	652864	28.4253	.00123762	848	719104	29 1204	.00117925
809	654481	28.4429	.00123609	849	720801	29.1376	.00117786
810	656100	28.4605	.00123457	850	722500	29.1548	.00117647
811	657721	28.4781	.00123305	851	724201	29.1719	.00117509
812	659344	28.4956	.00123153	852	725904	29.1890	.00117371
813	660969	28.5132	.00123001	853	727609	29.2062	.00117233
814	662596	28.5307	.00122850	854	729316	29.2233	.00117096
815	664225	28 5482	.00122699	855	731025	29.2404	.00116959
816	665856	28.5657	.00122549	856	732736	29.2575	.00116822
817	667489	28.5832	.00122399	857	734449	29.2746	.00116686
818	669124	28.6007	.00122249	858	736164	29.2916	.00116550
819	670761	28.6182	00122100	859	737881	29.3087	.00116414
820	672400	28.6356	.00121951	860	739600	29.3258	.00116279
821	674041	28.6531	.00121803	861	741321	29 3428	.00116144
822	675684	28.6705	.00121655	862	743044	29.3598	.00116009
823	677329	28.6880	.00121507	863	744769	29.3769	.00115875
824	678976	28.7054	.00121359	864	746496	29.3939	.00115741
825	680625	28.7228	.00121212	865	748225	29.4109	.00115607
826	682276	28.7402	.00121065	866	749956	29.4279	.00115473
827	683929	28.7576	.00120919	867	751689	29.4449	.00115340
828	685584	28.7750	.00120773	868	753424	29.4618	.00115207
829	687241	28.7924	.00120627	869	755161	29.4788	.00115075
830	688900	28.8097	.00120482	870	756900	29.4958	.00114943
831	690561	28.8271	.00120337	871	758641	29.5127	.00114811
832	692224	28.8444	.00120192	872	760384	29.5296	.00114679
833	693889	28.8617	.00120048	873	762129	29.5466	.00114548
834	695556	28.8791	.00119904	874	763876	29.5635	.00114416
835	697225	28.8964	.00119760	875	765625	29.5804	00114286
836	698896	28.9137	.00119617	876	767376	29.5973	.00114155
837	700569	28 9310	.00119474	877	769129	29.6142	.00114025
838	702244	28 9482	.00119332	878	770884	29.6311	.00113895
839	703921	28 9655	.00119190	879	772641	29.6479	.00113766
840	705600	28.9828	.00119048	880	774400	29.6648	.00113636

* Portions of Table II have been reproduced from J. W. Dunlap and A K Kurtz. *Handbook of Statistical Nomographs, Tables, and Formulas*, World Book Company, New York (1932), by permission of the authors and publishers.

Table II. Table of Squares, Square Roots, and Reciprocals of Numbers from 1 to 1,000*—Continued

N	N^2	$\sqrt{N}$	$1/N$	N	N^2	$\sqrt{N}$	$1/N$
881	776161	29.6816	.00113507	921	848241	30 3480	.00108578
882	777924	29 6985	.00113379	922	850084	30 3645	00108460
883	779689	29.7153	.00113250	923	851929	30.3809	.00108342
884	781456	29 7321	.00113122	924	853776	30.3974	.00108225
885	783225	29.7489	.00112994	925	855625	30.4138	.00108108
886	784996	29.7658	.00112867	926	857476	30.4302	.00107991
887	786769	29.7825	.00112740	927	859329	30 4467	.00107875
888	788544	29.7993	.00112613	928	861184	30.4631	.00107759
889	790321	29.8161	.00112486	929	863041	30 4795	.00107643
890	792100	29.8329	.00112360	930	864900	30 4959	.00107527
891	793881	29.8496	.00112233	931	866761	30 5123	.00107411
892	795664	29.8664	.00112108	932	868624	30.5287	.00107296
893	797449	29.8831	.00111982	933	870489	30.5450	.00107181
894	799236	29.8998	.00111857	934	872356	30.5614	.00107066
895	801025	29.9166	.00111732	935	874225	30.5778	.00106952
896	802816	29.9333	.00111607	936	876096	30.5941	.00106838
897	804609	29.9500	.00111483	937	877969	30.6105	.00106724
898	806404	29.9666	.00111359	938	879844	30.6268	.00106610
899	808201	29.9833	.00111235	939	881721	30.6431	.00106496
900	810000	30.0000	.00111111	940	883600	30.6594	00106383
901	811801	30.0167	.00110988	941	885481	30.6757	.00106270
902	813604	30.0333	.00110865	942	887364	30 6920	.00106157
903	815409	30 0500	.00110742	943	889249	30.7083	.00106045
904	817216	30.0666	.00110619	944	891136	30.7246	00105932
905	819025	30 0832	00110497	945	893025	30 7409	00105820
906	820836	30.0998	.00110375	946	894916	30 7571	00105708
907	822649	30.1164	.00110254	947	896809	30.7734	.00105597
908	824464	30.1330	.00110132	948	898704	30.7896	.00105485
909	826281	30.1496	.00110011	949	900601	30.8058	00105374
910	828100	30.1662	.00109890	950	902500	30 8221	00105263
911	829921	30.1828	.00109769	951	904401	30 8383	.00105152
912	831744	30.1993	.00109649	952	906304	30.8545	.00105042
913	833569	30 2159	.00109529	953	908209	30.8707	.00104932
914	835396	30.2324	.00109409	954	910116	30.8869	.00104822
915	837225	30 2490	.00109290	955	912025	30.9031	.00104712
916	839056	30.2655	.00109170	956	913936	30.9192	.00104603
917	840889	30.2820	.00109051	957	915849	30.9354	.00104493
918	842724	30 2985	.00108932	958	917764	30.9516	.00104384
919	844561	30.3150	.00108814	959	919681	30 9677	.00104275
920	846400	30.3315	.00108696	960	921600	30.9839	.00104167

* Portions of Table II have been reproduced from J W. Dunlap and A. K. Kurtz. *Handbook of Statistical Nomographs, Tables, and Formulas*, World Book Company, New York (1932), by permission of the authors and publishers.

Table II. Table of Squares, Square Roots, and Reciprocals of Numbers from 1 to 1,000*—Concluded

N	N^2	$\sqrt{N}$	$1/N$	N	N^2	$\sqrt{N}$	$1/N$
961	923521	31.0000	.00104058	981	962361	31 3209	.00101937
962	925444	31.0161	.00103950	982	964324	31.3369	.00101833
963	927369	31.0322	.00103842	983	966289	31 3528	.00101729
964	929296	31.0483	.00103734	984	968256	31.3688	.00101626
965	931225	31.0644	.00103627	985	970225	31 3847	.00101523
966	933156	31.0805	.00103520	986	972196	31 4006	.00101420
967	935089	31.0966	.00103413	987	974169	31 4166	.00101317
968	937024	31.1127	.00103306	988	976144	31.4325	.00101215
969	938961	31 1288	.00103199	989	978121	31.4484	.00101112
970	940900	31.1448	.00103093	990	980100	31.4643	.00101010
971	942841	31 1609	.00102987	991	982081	31.4802	.00100908
972	944784	31.1769	.00102881	992	984064	31.4960	.00100806
973	946729	31.1929	.00102775	993	986049	31.5119	.00100705
974	948676	31.2090	.00102669	994	988036	31.5278	.00100604
975	950625	31.2250	.00102564	995	990025	31 5436	.00100503
976	952576	31.2410	.00102459	996	992016	31.5595	.00100402
977	954529	31.2570	.00102354	997	994009	31.5753	00100301
978	956484	31 2730	.00102249	998	996004	31.5911	.00100200
979	958441	31.2890	.00102145	999	998001	31.6070	.00100100
980	960400	31.3050	.00102041	1000	1000000	31.6228	.00100000

* Portions of Table II have been reproduced from J. W. Dunlap and A. K. Kurtz. *Handbook of Statistical Nomographs, Tables, and Formulas*, World Book Company, New York (1932), by permission of the authors and publishers.

Table III. Areas and Ordinates of the Normal Curve in Terms of x/σ

(1) z Standard Score $\left(\frac{x}{\sigma}\right)$	(2) A Area from Mean to $\frac{x}{\sigma}$	(3) B Area in Larger Portion	(4) C Area in Smaller Portion	(5) y Ordinate at $\frac{x}{\sigma}$
0.00	.0000	.5000	.5000	.3989
0.01	.0040	.5040	.4960	.3989
0.02	.0080	.5080	.4920	.3989
0.03	.0120	.5120	.4880	.3988
0.04	.0160	.5160	.4840	.3986
0.05	.0199	.5199	.4801	.3984
0.06	.0239	.5239	.4761	.3982
0.07	.0279	.5279	.4721	.3980
0.08	.0319	.5319	.4681	.3977
0.09	.0359	.5359	.4641	.3973
0.10	.0398	.5398	.4602	.3970
0.11	.0438	.5438	.4562	.3965
0.12	.0478	.5478	.4522	.3961
0.13	.0517	.5517	.4483	.3956
0.14	.0557	.5557	.4443	.3951
0.15	.0596	.5596	.4404	.3945
0.16	.0636	.5636	.4364	.3939
0.17	.0675	.5675	.4325	.3932
0.18	.0714	.5714	.4286	.3925
0.19	.0753	.5753	.4247	.3918
0.20	.0793	.5793	.4207	.3910
0.21	.0832	.5832	.4168	.3902
0.22	.0871	.5871	.4129	.3894
0.23	.0910	.5910	.4090	.3885
0.24	.0948	.5948	.4052	.3876
0.25	.0987	.5987	.4013	.3867
0.26	.1026	.6026	.3974	.3857
0.27	.1064	.6064	.3936	.3847
0.28	.1103	.6103	.3897	.3836
0.29	.1141	.6141	.3859	.3825
0.30	.1179	.6179	.3821	.3814
0.31	.1217	.6217	.3783	.3802
0.32	.1255	.6255	.3745	.3790
0.33	.1293	.6293	.3707	.3778
0.34	.1331	.6331	.3669	.3765

Table III. Areas and Ordinates of the Normal Curve in Terms of x/σ—Continued

(1) z Standard Score $\left(\frac{x}{\sigma}\right)$	(2) A Area from Mean to $\frac{x}{\sigma}$	(3) B Area in Larger Portion	(4) C Area in Smaller Portion	(5) y Ordinate at $\frac{x}{\sigma}$
0.35	.1368	.6368	.3632	.3752
0.36	.1406	.6406	.3594	.3739
0.37	.1443	.6443	.3557	.3725
0.38	.1480	.6480	.3520	.3712
0.39	.1517	.6517	.3483	.3697
0.40	.1554	.6554	.3446	.3683
0.41	.1591	.6591	.3409	.3668
0.42	.1628	.6628	.3372	.3653
0.43	.1664	.6664	.3336	.3637
0.44	.1700	.6700	.3300	.3621
0.45	.1736	.6736	.3264	.3605
0.46	.1772	.6772	.3228	.3589
0.47	.1808	.6808	.3192	.3572
0.48	.1844	.6844	.3156	.3555
0.49	.1879	.6879	.3121	.3538
0.50	.1915	.6915	.3085	.3521
0.51	.1950	.6950	.3050	.3503
0.52	.1985	.6985	.3015	.3485
0.53	.2019	.7019	.2981	.3467
0.54	.2054	.7054	.2946	.3448
0.55	.2088	.7088	.2912	.3429
0.56	.2123	.7123	.2877	.3410
0.57	.2157	.7157	.2843	.3391
0.58	.2190	.7190	.2810	.3372
0.59	.2224	.7224	.2776	.3352
0.60	.2257	.7257	.2743	.3332
0.61	.2291	.7291	.2709	.3312
0.62	.2324	.7324	.2676	.3292
0.63	.2357	.7357	.2643	.3271
0.64	.2389	.7389	.2611	.3251
0.65	.2422	.7422	.2578	.3230
0.66	.2454	.7454	.2546	.3209
0.67	.2486	.7486	.2514	.3187
0.68	.2517	.7517	.2483	.3166
0.69	.2549	.7549	.2451	.3144

Table III. Areas and Ordinates of the Normal Curve in Terms of x/σ—Continued

(1) z Standard Score $\left(\frac{x}{\sigma}\right)$	(2) A Area from Mean to $\frac{x}{\sigma}$	(3) B Area in Larger Portion	(4) C Area in Smaller Portion	(5) y Ordinate at $\frac{x}{\sigma}$
0.70	.2580	.7580	.2420	.3123
0.71	.2611	.7611	.2389	.3101
0.72	.2642	.7642	.2358	.3079
0.73	.2673	.7673	.2327	.3056
0.74	.2704	.7704	.2296	.3034
0.75	.2734	.7734	.2266	.3011
0.76	.2764	.7764	.2236	.2989
0.77	.2794	.7794	.2206	.2966
0.78	.2823	.7823	.2177	.2943
0.79	.2852	.7852	.2148	.2920
0.80	.2881	.7881	.2119	.2897
0.81	.2910	.7910	.2090	.2874
0.82	.2939	.7939	.2061	.2850
0.83	.2967	.7967	.2033	.2827
0.84	.2995	.7995	.2005	.2803
0.85	.3023	.8023	.1977	.2780
0.86	.3051	.8051	.1949	.2756
0.87	.3078	.8078	.1922	.2732
0.88	.3106	.8106	.1894	.2709
0.89	.3133	.8133	.1867	.2685
0.90	.3159	.8159	.1841	.2661
0.91	.3186	.8186	.1814	.2637
0.92	.3212	.8212	.1788	.2613
0.93	.3238	.8238	.1762	.2589
0.94	.3264	.8264	.1736	.2565
0.95	.3289	.8289	.1711	.2541
0.96	.3315	.8315	.1685	.2516
0.97	.3340	.8340	.1660	.2492
0.98	.3365	.8365	.1635	.2468
0.99	.3389	.8389	.1611	.2444
1.00	.3413	.8413	.1587	.2420
1.01	.3438	.8438	.1562	.2396
1.02	.3461	.8461	.1539	.2371
1.03	.3485	.8485	.1515	.2347
1.04	.3508	.8508	.1492	.2323

Table III. Areas and Ordinates of the Normal Curve in Terms of x/σ—Continued

(1) z Standard Score $\left(\frac{x}{\sigma}\right)$	(2) A Area from Mean to $\frac{x}{\sigma}$	(3) B Area in Larger Portion	(4) C Area in Smaller Portion	(5) y Ordinate at $\frac{x}{\sigma}$
1.05	.3531	.8531	.1469	2299
1.06	.3554	.8554	.1446	2275
1.07	.3577	.8577	.1423	.2251
1.08	.3599	.8599	.1401	.2227
1.09	.3621	.8621	.1379	.2203
1.10	.3643	.8643	.1357	.2179
1.11	.3665	.8665	.1335	.2155
1.12	.3686	.8686	.1314	.2131
1.13	.3708	.8708	.1292	.2107
1.14	.3729	.8729	.1271	.2083
1.15	.3749	.8749	.1251	.2059
1.16	.3770	.8770	.1230	.2036
1.17	.3790	.8790	.1210	.2012
1.18	.3810	.8810	.1190	.1989
1.19	.3830	.8830	.1170	.1965
1.20	.3849	.8849	.1151	.1942
1.21	.3869	.8869	.1131	.1919
1.22	.3888	.8888	.1112	.1895
1.23	.3907	.8907	.1093	.1872
1.24	.3925	.8925	.1075	.1849
1.25	.3944	.8944	.1056	.1826
1.26	.3962	.8962	.1038	.1804
1.27	.3980	.8980	.1020	.1781
1.28	.3997	.8997	.1003	.1758
1.29	.4015	.9015	.0985	.1736
1.30	.4032	.9032	.0968	.1714
1.31	.4049	.9049	.0951	.1691
1.32	.4066	.9066	.0934	.1669
1.33	.4082	.9082	.0918	.1647
1.34	.4099	.9099	.0901	.1626
1.35	.4115	.9115	.0885	.1604
1.36	.4131	.9131	.0869	.1582
1.37	.4147	.9147	.0853	.1561
1.38	.4162	.9162	.0838	.1539
1.39	.4177	.9177	.0823	.1518

Table III. Areas and Ordinates of the Normal Curve in Terms of x/σ—Continued

(1) z Standard Score $\left(\frac{x}{\sigma}\right)$	(2) A Area from Mean to $\frac{x}{\sigma}$	(3) B Area in Larger Portion	(4) C Area in Smaller Portion	(5) y Ordinate at $\frac{x}{\sigma}$
1.40	.4192	.9192	.0808	.1497
1.41	.4207	.9207	.0793	.1476
1.42	.4222	.9222	.0778	.1456
1.43	.4236	.9236	.0764	.1435
1.44	.4251	.9251	.0749	.1415
1.45	.4265	.9265	.0735	.1394
1.46	.4279	.9279	.0721	.1374
1.47	.4292	.9292	.0708	.1354
1.48	.4306	.9306	.0694	.1334
1.49	.4319	.9319	.0681	.1315
1.50	.4332	.9332	.0668	.1295
1.51	.4345	.9345	.0655	.1276
1.52	.4357	.9357	.0643	.1257
1.53	.4370	.9370	.0630	.1238
1.54	.4382	.9382	.0618	.1219
1.55	.4394	.9394	.0606	.1200
1.56	.4406	.9406	.0594	.1182
1.57	.4418	.9418	.0582	.1163
1.58	.4429	.9429	.0571	.1145
1.59	.4441	.9441	.0559	.1127
1.60	.4452	.9452	.0548	.1109
1.61	.4463	.9463	.0537	.1092
1.62	.4474	.9474	.0526	.1074
1.63	.4484	.9484	.0516	.1057
1.64	.4495	.9495	.0505	.1040
1.65	.4505	.9505	.0495	.1023
1.66	.4515	.9515	.0485	.1006
1.67	.4525	.9525	.0475	.0989
1.68	.4535	.9535	.0465	.0973
1.69	.4545	.9545	.0455	.0957
1.70	.4554	.9554	.0446	.0940
1.71	.4564	.9564	.0436	.0925
1.72	.4573	.9573	.0427	.0909
1.73	.4582	.9582	.0418	.0893
1.74	.4591	.9591	.0409	.0878

Table III. Areas and Ordinates of the Normal Curve in Terms of x/σ—Continued

(1) z Standard Score $\left(\frac{x}{\sigma}\right)$	(2) A Area from Mean to $\frac{x}{\sigma}$	(3) B Area in Larger Portion	(4) C Area in Smaller Portion	(5) y Ordinate at $\frac{x}{\sigma}$
1.75	.4599	.9599	.0401	.0863
1.76	.4608	.9608	.0392	.0848
1.77	.4616	.9616	.0384	.0833
1.78	.4625	.9625	.0375	.0818
1.79	.4633	.9633	.0367	.0804
1.80	.4641	.9641	.0359	.0790
1.81	.4649	.9649	.0351	.0775
1.82	.4656	.9656	.0344	.0761
1.83	.4664	.9664	.0336	.0748
1.84	.4671	.9671	.0329	.0734
1.85	.4678	.9678	.0322	.0721
1.86	.4686	.9686	.0314	.0707
1.87	.4693	.9693	.0307	.0694
1.88	.4699	.9699	.0301	.0681
1.89	.4706	.9706	.0294	.0669
1.90	.4713	.9713	.0287	.0656
1.91	.4719	.9719	.0281	.0644
1.92	.4726	.9726	.0274	.0632
1.93	.4732	.9732	.0268	.0620
1.94	.4738	.9738	.0262	.0608
1.95	.4744	.9744	.0256	.0596
1.96	.4750	.9750	.0250	.0584
1.97	.4756	.9756	.0244	.0573
1.98	.4761	.9761	.0239	.0562
1.99	.4767	.9767	.0233	.0551
2.00	.4772	.9772	.0228	.0540
2.01	.4778	.9778	.0222	.0529
2.02	.4783	.9783	.0217	.0519
2.03	.4788	.9788	.0212	.0508
2.04	.4793	.9793	.0207	.0498
2.05	.4798	.9798	.0202	.0488
2.06	.4803	.9803	.0197	.0478
2.07	.4808	.9808	.0192	.0468
2.08	.4812	.9812	.0188	.0459
2.09	.4817	.9817	.0183	.0449

Table III. Areas and Ordinates of the Normal Curve in Terms of x/σ—Continued

(1) z Standard Score $\left(\frac{x}{\sigma}\right)$	(2) A Area from Mean to $\frac{x}{\sigma}$	(3) B Area in Larger Portion	(4) C Area in Smaller Portion	(5) y Ordinate at $\frac{x}{\sigma}$
2.10	.4821	.9821	.0179	.0440
2.11	.4826	.9826	.0174	.0431
2.12	.4830	.9830	.0170	.0422
2.13	.4834	.9834	.0166	.0413
2.14	.4838	.9838	.0162	.0404
2.15	.4842	.9842	.0158	.0396
2.16	.4846	.9846	.0154	.0387
2.17	.4850	.9850	.0150	.0379
2.18	.4854	.9854	.0146	.0371
2.19	.4857	.9857	.0143	.0363
2.20	.4861	.9861	.0139	.0355
2.21	.4864	.9864	.0136	.0347
2.22	.4868	.9868	.0132	.0339
2.23	.4871	.9871	.0129	.0332
2.24	.4875	.9875	.0125	.0325
2.25	.4878	.9878	.0122	.0317
2.26	.4881	.9881	.0119	.0310
2.27	.4884	.9884	.0116	.0303
2.28	.4887	.9887	.0113	.0297
2.29	.4890	.9890	.0110	.0290
2.30	.4893	.9893	.0107	.0283
2.31	.4896	.9896	.0104	.0277
2.32	.4898	.9898	.0102	.0270
2.33	.4901	.9901	.0099	.0264
2.34	.4904	.9904	.0096	.0258
2.35	.4906	.9906	.0094	.0252
2.36	.4909	.9909	.0091	.0246
2.37	.4911	.9911	.0089	.0241
2.38	.4913	.9913	.0087	.0235
2.39	.4916	.9916	.0084	.0229
2.40	.4918	.9918	.0082	.0224
2.41	.4920	.9920	.0080	.0219
2.42	.4922	.9922	.0078	.0213
2.43	.4925	.9925	.0075	.0208
2.44	.4927	.9927	.0073	.0203

Table III. Areas and Ordinates of the Normal Curve in Terms of x/σ—Continued

(1) z Standard Score $\left(\frac{x}{\sigma}\right)$	(2) A Area from Mean to $\frac{x}{\sigma}$	(3) B Area in Larger Portion	(4) C Area in Smaller Portion	(5) y Ordinate at $\frac{x}{\sigma}$
2.45	.4929	.9929	.0071	.0198
2.46	.4931	.9931	.0069	.0194
2.47	.4932	.9932	.0068	.0189
2.48	.4934	.9934	.0066	.0184
2.49	.4936	.9936	.0064	.0180
2.50	.4938	.9938	.0062	.0175
2.51	.4940	.9940	.0060	.0171
2.52	.4941	.9941	.0059	.0167
2.53	.4943	.9943	.0057	.0163
2.54	.4945	.9945	.0055	.0158
2.55	.4946	.9946	.0054	.0154
2.56	.4948	.9948	.0052	.0151
2.57	.4949	.9949	.0051	.0147
2.58	.4951	.9951	.0049	.0143
2.59	.4952	.9952	.0048	.0139
2.60	.4953	.9953	.0047	.0136
2.61	.4955	.9955	.0045	.0132
2.62	.4956	.9956	.0044	.0129
2.63	.4957	.9957	.0043	.0126
2.64	.4959	.9959	.0041	.0122
2.65	.4960	.9960	.0040	.0119
2.66	.4961	.9961	.0039	.0116
2.67	.4962	.9962	.0038	.0113
2.68	.4963	.9963	.0037	.0110
2.69	.4964	.9964	.0036	.0107
2.70	.4965	.9965	.0035	.0104
2.71	.4966	.9966	.0034	.0101
2.72	.4967	.9967	.0033	.0099
2.73	.4968	.9968	.0032	.0096
2.74	.4969	.9969	.0031	.0093
2.75	.4970	.9970	.0030	.0091
2.76	.4971	.9971	.0029	.0088
2.77	.4972	.9972	.0028	.0086
2.78	.4973	.9973	.0027	.0084
2.79	.4974	.9974	.0026	.0081

Table III. Areas and Ordinates of the Normal Curve in Terms of x/σ—Continued

(1) z Standard Score $\left(\frac{x}{\sigma}\right)$	(2) A Area from Mean to $\frac{x}{\sigma}$	(3) B Area in Larger Portion	(4) C Area in Smaller Portion	(5) y Ordinate at $\frac{x}{\sigma}$
2.80	.4974	.9974	.0026	.0079
2.81	.4975	.9975	.0025	.0077
2.82	.4976	.9976	.0024	.0075
2.83	.4977	.9977	.0023	.0073
2.84	.4977	.9977	.0023	.0071
2.85	.4978	.9978	.0022	.0069
2.86	.4979	.9979	.0021	.0067
2.87	.4979	.9979	.0021	.0065
2.88	.4980	.9980	.0020	.0063
2.89	.4981	.9981	.0019	.0061
2.90	.4981	.9981	.0019	.0060
2.91	.4982	.9982	.0018	.0058
2.92	.4982	.9982	.0018	.0056
2.93	.4983	.9983	.0017	.0055
2.94	.4984	.9984	.0016	.0053
2.95	.4984	.9984	.0016	.0051
2.96	.4985	.9985	.0015	.0050
2.97	.4985	.9985	.0015	.0048
2.98	.4986	.9986	0014	.0047
2.99	.4986	.9986	.0014	.0046
3.00	.4987	.9987	.0013	.0044
3.01	.4987	.9987	.0013	.0043
3.02	.4987	.9987	.0013	.0042
3.03	.4988	.9988	.0012	.0040
3.04	.4988	.9988	.0012	.0039
3.05	.4989	.9989	.0011	.0038
3.06	.4989	.9989	.0011	.0037
3.07	.4989	.9989	.0011	.0036
3.08	.4990	.9990	.0010	.0035
3.09	.4990	.9990	.0010	.0034
3.10	.4990	.9990	.0010	.0033
3.11	.4991	.9991	.0009	.0032
3.12	.4991	.9991	.0009	.0031
3.13	.4991	.9991	.0009	.0030
3.14	.4992	.9992	.0008	.0029

Table III. Areas and Ordinates of the Normal Curve in Terms of x/σ—Concluded

(1) z Standard Score $\left(\frac{x}{\sigma}\right)$	(2) A Area from Mean to $\frac{x}{\sigma}$	(3) B Area in Larger Portion	(4) C Area in Smaller Portion	(5) y Ordinate at $\frac{x}{\sigma}$
3.15	.4992	.9992	.0008	.0028
3.16	.4992	.9992	.0008	.0027
3.17	.4992	.9992	.0008	.0026
3.18	.4993	.9993	.0007	.0025
3.19	.4993	.9993	.0007	.0025
3.20	.4993	.9993	.0007	.0024
3.21	.4993	.9993	.0007	.0023
3.22	.4994	.9994	.0006	.0022
3.23	.4994	.9994	.0006	.0022
3.24	.4994	.9994	.0006	.0021
3.30	.4995	.9995	.0005	.0017
3.40	.4997	.9997	.0003	.0012
3.50	.4998	.9998	.0002	.0009
3.60	.4998	.9998	.0002	.0006
3.70	.4999	.9999	.0001	.0004

Table IV. Table of χ^{2*}

Degrees of Freedom *df*	$P = .99$	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01
1	.000157	.000628	.00393	.0158	.0642	.148	.455	1.074	1 642	2.706	3 841	5 412	6 635
2	.0201	.0404	103	.211	.446	713	1.386	2 408	3.219	4.605	5 991	7.824	9 210
3	.115	.185	352	584	1 005	1 424	2 366	3 665	4.642	6 251	7.815	9.837	11 341
4	.297	.429	711	1 064	1 649	2 195	3.357	4 878	5 989	7 779	9 488	11 668	13 277
5	.554	.752	1.145	1 610	2 343	3 000	4 351	6 064	7.289	9 236	11.070	13 388	15 086
6	.872	1.134	1 635	2 204	3.070	3 828	5 348	7.231	8.558	10 645	12.592	15 033	16.812
7	1.239	1 564	2 167	2.833	3.822	4.671	6.346	8.383	9.803	12 017	14 067	16 622	18 475
8	1.646	2 032	2.733	3 490	4 594	5.527	7 344	9 524	11.030	13 362	15 507	18 168	20.090
9	2 088	2 532	3 325	4 168	5.380	6 393	8.343	10 656	12 242	14 684	16 919	19 679	21 666
10	2 558	3 059	3 940	4 865	6 179	7 267	9.342	11.781	13.442	15 987	18 307	21 161	23 209
11	3 053	3 609	4 575	5.578	6 989	8 148	10 341	12 899	14 631	17 275	19 675	22 618	24 725
12	3 571	4.178	5 226	6 304	7.807	9 034	11 340	14 011	15.812	18.549	21.026	24 054	26 217
13	4.107	4.765	5 892	7 042	8 634	9 926	12 340	15 119	16 985	19.812	22 362	25 472	27 688
14	4 660	5 368	6 571	7.790	9.467	10 821	13.339	16 222	18 151	21 064	23 685	26 873	29.141
15	5 229	5 985	7.261	8 547	10 307	11 721	14.339	17.322	19 311	22 307	24 996	28 259	30 578
16	5 812	6 614	7 962	9 312	11 152	12 624	15 338	18 418	20.465	23 542	26 296	29 633	32 000
17	6.408	7 255	8 672	10.085	12 002	13 531	16 338	19 511	21 615	24 769	27 587	30.995	33.409
18	7.015	7.906	9.390	10.865	12 857	14 440	17 338	20.601	22 760	25 989	28 869	32 346	34 805
19	7.633	8.567	10.117	11 651	13.716	15 352	18 338	21 689	23 900	27 204	30.144	33 687	36.191
20	8.260	9 237	10 851	12 443	14.578	16 266	19.337	22 775	25 038	28.412	31 410	35 020	37.566
21	8.897	9 915	11.591	13 240	15 445	17 182	20 337	23 858	26.171	29 615	32 671	36 343	38 932
22	9 542	10 600	12 338	14.041	16 314	18 101	21 337	24 939	27 301	30 813	33 924	37 659	40 289
23	10 196	11 293	13 091	14.848	17 187	19 021	22.337	26.018	28 429	32 007	35 172	38 968	41 638
24	10 856	11 992	13.848	15 659	18 062	19 943	23.337	27.096	29.553	33 196	36 415	40 270	42.980
25	11.524	12 697	14.611	16 473	18.940	20.867	24 337	28 172	30 675	34 382	37 652	41.566	44 314
26	12 198	13 409	15 379	17 292	19.820	21 792	25.336	29 246	31 795	35 563	38.885	42 856	45 642
27	12.879	14 125	16 151	18.114	20.703	22 719	26 336	30 319	32.912	36 741	40.113	44 140	46 963
28	13 565	14 847	16.928	18.939	21.588	23 647	27.336	31 391	34 027	37 916	41 337	45 419	48 278
29	14 256	15.574	17 708	19.768	22 475	24.577	28 336	32 461	35.139	39 087	42 557	46 693	49 588
30	14.953	16.306	18 493	20 599	23 364	25 508	29.336	33 530	36.250	40 256	43.773	47.962	50.892

* Table IV is reprinted from Table III of Fisher· *Statistical Methods for Research Workers*, Oliver & Boyd Ltd., Edinburgh, by permission of the author and publishers.

For larger values of *df*, the expression $\sqrt{2\chi^2} - \sqrt{2(df) - 1}$ may be used as a normal deviate with unit standard error.

Table V. Table of t*

df	P = .450	.400	.350	.300	.250	.200	.150	.100	.050	.025	.010	.005
1	.158	.325	.510	.727	1 000	1.376	1 963	3 078	6.314	12 706	31.821	63 657
2	.142	.289	445	.617	.816	1 061	1 386	1 886	2 920	4 303	6 965	9 925
3	.137	277	.424	.584	765	978	1.250	1.638	2.353	3 182	4.541	5 841
4	.134	.271	.414	.569	.741	941	1 190	1 533	2 132	2 776	3 747	4.604
5	.132	267	.408	.559	.727	.920	1 156	1.476	2.015	2.571	3 365	4.032
6	.131	265	.404	.553	.718	.906	1 134	1 440	1.943	2 447	3.143	3.707
7	.130	.263	.402	.549	.711	.896	1.119	1 415	1 895	2.365	2 998	3.499
8	.130	.262	.399	.546	.706	.889	1 108	1.397	1.860	2.306	2.896	3.355
9	.129	.261	.398	.543	.703	.883	1.100	1 383	1.833	2.262	2.821	3 250
10	129	.260	.397	.542	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169
11	.129	.260	.396	.540	.697	.876	1.088	1.363	1.796	2.201	2.718	3.106
12	.128	.259	.395	.539	.695	.873	1.083	1.356	1.782	2.179	2.681	3.055
13	.128	.259	.394	.538	.694	870	1.079	1.350	1.771	2.160	2.650	3.012
14	.128	.258	.393	.537	.692	.868	1.076	1.345	1.761	2.145	2.624	2.977
15	.128	.258	.393	.536	.691	.866	1.074	1.341	1.753	2.131	2.602	2 947
16	.128	.258	.392	.535	.690	.865	1.071	1.337	1 746	2 120	2.583	2.921
17	.128	.257	.392	.534	.689	863	1.069	1 333	1.740	2.110	2.567	2.898
18	.127	.257	.392	.534	.688	.862	1.067	1.330	1.734	2.101	2.552	2.878
19	.127	.257	.391	.533	.688	861	1.066	1.328	1.729	2.093	2 539	2.861
20	.127	.257	.391	.533	.687	.860	1.064	1 325	1.725	2.086	2 528	2.845
21	.127	.257	.391	.532	.686	.859	1.063	1 323	1.721	2.080	2.518	2.831
22	.127	.256	.390	.532	.686	.858	1 061	1 321	1.717	2.074	2.508	2.819
23	.127	.256	.390	532	685	858	1.060	1.319	1.714	2.069	2.500	2.807
24	.127	.256	.390	.531	.685	.857	1 059	1.318	1.711	2.064	2.492	2.797
25	.127	.256	.390	.531	.684	.856	1.058	1.316	1.708	2.060	2.485	2.787
26	.127	.256	.390	531	.684	.856	1.058	1.315	1.706	2.056	2.479	2.779
27	.127	.256	.389	.531	.684	.855	1.057	1.314	1.703	2.052	2.473	2 771
28	.127	.256	.389	530	.683	.855	1.056	1.313	1 701	2.048	2.467	2 763
29	.127	.256	.389	.530	.683	.854	1.055	1.311	1 699	2 045	2.462	2.756
30	.127	256	.389	530	.683	.854	1.055	1.310	1.697	2 042	2.457	2 750
∞	.12566	.25335	.38532	.52440	.67449	.84162	1.03643	1.28155	1.64485	1.95996	2.32634	2 57582

Additional Values of t at the .025 and .005 Levels of Significance†

df	.025	.005	df	.025	005	df	025	.005
32	2.037	2 739	55	2 005	2.668	125	1 979	2 616
34	2.032	2.728	60	2.000	2 660	150	1 976	2.609
36	2.027	2.718	65	1.998	2.653	175	1 974	2.605
38	2.025	2.711	70	1.994	2.648	200	1.972	2.601
40	2 021	2.704	75	1 992	2.643	300	1.968	2.592
42	2.017	2.696	80	1 990	2 638	400	1 966	2.588
44	2.015	2.691	85	1 989	2.635	500	1.965	2.586
46	2.012	2.685	90	1 987	2 632	1000	1.962	2.581
48	2.010	2.681	95	1.986	2 629	∞	1 960	2 576
50	2.008	2.678	100	1.984	2.626			

* Table V is reprinted from Table IV of Fisher: *Statistical Methods for Research Workers*, Oliver & Boyd Ltd., Edinburgh, by permission of the author and publishers.

† Additional entries were taken from Snedecor: *Statistical Methods*, Iowa State College Press, Ames, Iowa, by permission of the author and publisher. Values for 75, 85, 95, and 175 degrees of freedom were obtained by linear interpolation.

The probabilities given are for a one-sided test.

Table VI. Values of the Correlation Coefficient for Different Levels of Significance*

df	*P* = .050	.025	.010	.005
1	988	.997	.9995	.9999
2	900	950	.980	.990
3	805	.878	.934	.959
4	729	.811	.882	.917
5	.669	.754	.833	.874
6	622	707	.789	.834
7	582	.666	.750	.798
8	549	.632	.716	.765
9	.521	.602	.685	.735
10	497	.576	.658	.708
11	.476	553	.634	.684
12	458	.532	.612	.661
13	.441	.514	592	.641
14	.426	.497	.574	.623
15	.412	.482	558	.606
16	.400	.468	.542	.590
17	.389	.456	.528	.575
18	.378	.444	516	561
19	369	.433	.503	.549
20	.360	.423	.492	.537
21	352	.413	.482	.526
22	.344	.404	.472	.515
23	.337	.396	.462	.505
24	.330	.388	.453	.496
25	.323	.381	.445	.487
26	.317	.374	.437	.479
27	.311	.367	.430	.471
28	.306	.361	.423	.463
29	301	.355	.416	.456
30	.296	.349	.409	.449
35	.275	.325	.381	.418
40	.257	304	.358	393
45	243	.288	338	.372
50	.231	273	.322	.354
60	.211	.250	.295	.325
70	.195	.232	.274	.302
80	.183	.217	.256	.283
90	.173	.205	.242	.267
100	.164	195	.230	254

Additional values of *r* at the .025 and .005 Levels of Significance

df	025	.005	*df*	025	.005	*df*	025	005
32	339	436	48	279	.361	150	159	.208
34	329	.424	55	.261	338	175	.148	.193
36	.320	.413	65	.241	.313	200	.138	.181
38	.312	.403	75	.224	.292	300	.113	.148
42	.297	.384	85	.211	.275	400	098	.128
44	.291	.376	95	.200	.260	500	.088	.115
46	.284	.368	125	.174	.228	1,000	.062	081

*Table VI is reprinted from Table V.A. of R. A. Fisher, *Statistical Methods for Research Workers*, Oliver & Boyd Ltd., Edinburgh, by permission of the author and publishers.

Additional entries were calculated using the table of *t*.

The probabilities given are for a one-sided test.

Table VII. Table of z' Values for r*

r	z'	r	z'	r	z'	r	z'	r	z'
.000	.000	.200	.203	.400	.424	.600	.693	.800	1.099
.005	.005	.205	.208	.405	.430	.605	.701	.805	1.113
.010	.010	.210	.213	.410	.436	.610	.709	.810	1.127
.015	.015	.215	.218	.415	.442	.615	.717	.815	1.142
.020	.020	.220	.224	.420	.448	.620	.725	.820	1.157
.025	.025	.225	.229	.425	.454	.625	.733	.825	1.172
.030	.030	.230	.234	.430	.460	.630	.741	.830	1.188
.035	.035	.235	.239	.435	.466	.635	.750	.835	1.204
.040	.040	.240	.245	.440	.472	.640	.758	.840	1.221
.045	.045	.245	.250	.445	.478	.645	.767	.845	1.238
.050	.050	.250	.255	.450	.485	.650	.775	.850	1.256
.055	.055	.255	.261	.455	.491	.655	.784	.855	1.274
.060	.060	.260	.266	.460	.497	.660	.793	.860	1.293
.065	.065	.265	.271	.465	.504	.665	.802	.865	1.313
.070	.070	.270	.277	.470	.510	.670	.811	.870	1.333
.075	.075	.275	.282	.475	.517	.675	.820	.875	1.354
.080	.080	.280	.288	.480	.523	.680	.829	.880	1.376
.085	.085	.285	.293	.485	.530	.685	.838	.885	1.398
.090	.090	.290	.299	.490	.536	.690	.848	.890	1.422
.095	.095	.295	.304	.495	.543	.695	.858	.895	1.447
.100	.100	.300	.310	.500	.549	.700	.867	.900	1.472
.105	.105	.305	.315	.505	.556	.705	.877	.905	1.499
.110	.110	.310	.321	.510	.563	.710	.887	.910	1.528
.115	.116	.315	.326	.515	.570	.715	.897	.915	1.557
.120	.121	.320	.332	.520	.576	.720	.908	.920	1.589
.125	.126	.325	.337	.525	.583	.725	.918	.925	1.623
.130	.131	.330	.343	.530	.590	.730	.929	.930	1.658
.135	.136	.335	.348	.535	.597	.735	.940	.935	1.697
.140	.141	.340	.354	.540	.604	.740	.950	.940	1.738
.145	.146	.345	.360	.545	.611	.745	.962	.945	1.783
.150	.151	.350	.365	.550	.618	.750	.973	.950	1.832
.155	.156	.355	.371	.555	.626	.755	.984	.955	1.886
.160	.161	.360	.377	.560	.633	.760	.996	.960	1.946
.165	.167	.365	.383	.565	.640	.765	1.008	.965	2 014
.170	.172	.370	.388	.570	.648	.770	1.020	.970	2.092
.175	.177	.375	.394	.575	.655	.775	1.033	.975	2.185
.180	.182	.380	.400	.580	.662	.780	1.045	980	2.298
.185	.187	.385	.406	.585	.670	.785	1.058	.985	2 443
.190	.192	.390	.412	.590	.678	.790	1.071	.990	2.647
.195	.198	.395	.418	.595	.685	.795	1.085	.995	2.994

*Table VII was constructed by F. P. Kilpatrick and D A Buchanan from formula (6 3)

Table VIII. The 5 (Roman Type) and 1 (Boldface Type) Per Cent Points for the Distribution of F^*

n_2	n_1 degrees of freedom (for greater mean square)																						
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
1	161 **4,052**	200 **4,999**	216 **5,403**	225 **5,625**	230 **5,764**	234 **5,859**	237 **5,928**	239 **5,981**	241 **6,022**	242 **6,056**	243 **6,082**	244 **6,106**	245 **6,142**	246 **6,169**	248 **6,208**	249 **6,234**	250 **6,258**	251 **6,286**	252 **6,302**	253 **6,323**	253 **6,334**	254 **6,352**	254 **6,361**	254 **6,366**
2	18 51 **98.49**	19 00 **99.00**	19.16 **99.17**	19.25 **99.25**	19 30 **99.30**	19.33 **99.33**	19.36 **99.34**	19 37 **99.36**	19 38 **99.38**	19 39 **99.40**	19.40 **99.41**	19 41 **99.42**	19.42 **99.43**	19 43 **99.44**	19 44 **99.45**	19 45 **99.46**	19 46 **99.47**	19.47 **99.48**	19.47 **99.48**	19.48 **99.49**	19.49 **99.49**	19 49 **99.49**	19 50 **99.50**	19 50 **99.50**
3	10 13 **34.12**	9.55 **30.82**	9.28 **29.46**	9.12 **28.71**	9 01 **28.24**	8 94 **27.91**	8.88 **27.67**	8 84 **27.49**	8 81 **27.34**	8 78 **27.23**	8 76 **27.13**	8 74 **27.05**	8.71 **26.92**	8 69 **26.83**	8 66 **26.69**	8.64 **26.60**	8 62 **26.50**	8.60 **26.41**	8 58 **26.35**	8 57 **26.27**	8 56 **26.23**	8 54 **26.18**	8 54 **26.14**	8.53 **26.12**
4	7 71 **21.20**	6 94 **18.00**	6 59 **16.69**	6.39 **15.98**	6 26 **15.52**	6.16 **15.21**	6.09 **14.98**	6 04 **14.80**	6 00 **14.66**	5 96 **14.54**	5 93 **14.45**	5 91 **14.37**	5 87 **14.24**	5 84 **14.15**	5 80 **14.02**	5 77 **13.93**	5 74 **13.83**	5 71 **13.74**	5.70 **13.69**	5 68 **13.61**	5 66 **13.57**	5 65 **13.52**	5 64 **13.48**	5 63 **13.46**
5	6 61 **16.26**	5 79 **13.27**	5 41 **12.06**	5.19 **11.39**	5.05 **10.97**	4.95 **10.67**	4.88 **10.45**	4 82 **10.27**	4.78 **10.15**	4.74 **10.05**	4.70 **9.96**	4.68 **9.89**	4 64 **9.77**	4 60 **9.68**	4 56 **9.55**	4 53 **9.47**	4 50 **9.38**	4 46 **9.29**	4 44 **9.24**	4 42 **9.17**	4 40 **9.13**	4 38 **9.07**	4.37 **9.04**	4.36 **9.02**
6	5.99 **13.74**	5.14 **10.92**	4 76 **9.78**	4 53 **9.15**	4 39 **8.75**	4 28 **8.47**	4.21 **8.26**	4 15 **8.10**	4 10 **7.98**	4 06 **7.87**	4 03 **7.79**	4 00 **7.72**	3 96 **7.60**	3 92 **7.52**	3 87 **7.39**	3 84 **7.31**	3 81 **7.23**	3.77 **7.14**	3 75 **7.09**	3 72 **7.02**	3 71 **6.99**	3.69 **6.94**	3.68 **6.90**	3 67 **6.88**
7	5 59 **12.25**	4 74 **9.55**	4 35 **8.45**	4 12 **7.85**	3 97 **7.46**	3 87 **7.19**	3 79 **7.00**	3 73 **6.84**	3.68 **6.71**	3 63 **6.62**	3 60 **6.54**	3 57 **6.47**	3 52 **6.35**	3 49 **6.27**	3 44 **6.15**	3 41 **6.07**	3.38 **5.98**	3 34 **5.90**	3 32 **5.85**	3 29 **5.78**	3 28 **5.75**	3 25 **5.70**	3.24 **5.67**	3.23 **5.65**
8	5.32 **11.26**	4 46 **8.65**	4 07 **7.59**	3 84 **7.01**	3 69 **6.63**	3 58 **6.37**	3.50 **6.19**	3.44 **6.03**	3 39 **5.91**	3.34 **5.82**	3.31 **5.74**	3 28 **5.67**	3.23 **5.56**	3 20 **5.48**	3 15 **5.36**	3 12 **5.28**	3 08 **5.20**	3.05 **5.11**	3 03 **5.06**	3.00 **5.00**	2 98 **4.96**	2 96 **4.91**	2 94 **4.88**	2 93 **4.86**
9	5.12 **10.56**	4 26 **8.02**	3.86 **6.99**	3.63 **6.42**	3 48 **6.06**	3 37 **5.80**	3 29 **5.62**	3 23 **5.47**	3 18 **5.35**	3.13 **5.26**	3 10 **5.18**	3 07 **5.11**	3 02 **5.00**	2 98 **4.92**	2 93 **4.80**	2 90 **4.73**	2.86 **4.64**	2 82 **4.56**	2 80 **4.51**	2 77 **4.45**	2 76 **4.41**	2 73 **4.36**	2 72 **4.33**	2 71 **4.31**
10	4 96 **10.04**	4 10 **7.56**	3 71 **6.55**	3.48 **5.99**	3 33 **5.64**	3 22 **5.39**	3.14 **5.21**	3.07 **5.06**	3 02 **4.95**	2.97 **4.85**	2 94 **4.78**	2 91 **4.71**	2 86 **4.60**	2 82 **4.52**	2 77 **4.41**	2.74 **4.33**	2 70 **4.25**	2 67 **4.17**	2 64 **4.12**	2 61 **4.05**	2 59 **4.01**	2 56 **3.96**	2.55 **3.93**	2.54 **3.91**
11	4 84 **9.65**	3 98 **7.20**	3 59 **6.22**	3 36 **5.67**	3 20 **5.32**	3 09 **5.07**	3 01 **4.88**	2 95 **4.74**	2 90 **4.63**	2 86 **4.54**	2 82 **4.46**	2 79 **4.40**	2 74 **4.29**	2 70 **4.21**	2 65 **4.10**	2 61 **4.02**	2 57 **3.94**	2 53 **3.86**	2 50 **3.80**	2 47 **3.74**	2 45 **3.70**	2 42 **3.66**	2 41 **3.62**	2 40 **3.60**
12	4.75 **9.33**	3.88 **6.93**	3 49 **5.95**	3 26 **5.41**	3.11 **5.06**	3.00 **4.82**	2 92 **4.65**	2 85 **4.50**	2.80 **4.39**	2 76 **4.30**	2.72 **4.22**	2 69 **4.16**	2 64 **4.05**	2 60 **3.98**	2 54 **3.86**	2.50 **3.78**	2 46 **3.70**	2 42 **3.61**	2 40 **3.56**	2 36 **3.49**	2 35 **3.46**	2 32 **3.41**	2 31 **3.38**	2 30 **3.36**
13	4.67 **9.07**	3 80 **6.70**	3 41 **5.74**	3.18 **5.20**	3 02 **4.86**	2 92 **4.62**	2 84 **4.44**	2 77 **4.30**	2 72 **4.19**	2 67 **4.10**	2 63 **4.02**	2 60 **3.96**	2 55 **3.85**	2 51 **3.78**	2 46 **3.67**	2.42 **3.59**	2 38 **3.51**	2 34 **3.42**	2 32 **3.37**	2 28 **3.30**	2 26 **3.27**	2.24 **3.21**	2.22 **3.18**	2.21 **3.16**

* Table VIII is reproduced from Snedecor· *Statistical Methods*, Iowa State College Press, Ames, Iowa, by permission of the author and publisher.

Table VIII. The 5 (Roman Type) and 1 (Boldface Type) Per Cent Points for the Distribution of F*—Continued

n_2	n_1 degrees of freedom (for greater mean square)																							
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
14	4 60 **8.86**	3 74 **6.51**	3 34 **5.56**	3 11 **5.03**	2 96 **4.69**	2 85 **4.46**	2 77 **4.28**	2 70 **4.14**	2 65 **4.03**	2 60 **3.94**	2 56 **3.86**	2 53 **3.80**	2 48 **3.70**	2 44 **3.62**	2 39 **3.51**	2 35 **3.43**	2 31 **3.34**	2 27 **3.26**	2 24 **3.21**	2 21 **3.14**	2 19 **3.11**	2 16 **3.06**	2 14 **3.02**	2.13 **3.00**
15	4 54 **8.68**	3 68 **6.36**	3 29 **5.42**	3 06 **4.89**	2 90 **4.56**	2 79 **4.32**	2 70 **4.14**	2 64 **4.00**	2.59 **3.89**	2 55 **3.80**	2.51 **3.73**	2 48 **3.67**	2 43 **3.56**	2 39 **3.48**	2.33 **3.36**	2 29 **3.29**	2 25 **3.20**	2 21 **3.12**	2 18 **3.07**	2.15 **3.00**	2.12 **2.97**	2.10 **2.92**	2 08 **2.89**	2 07 **2.87**
16	4.49 **8.53**	3.63 **6.23**	3.24 **5.29**	3.01 **4.77**	2 85 **4.44**	2 74 **4.20**	2 66 **4.03**	2 59 **3.89**	2 54 **3.78**	2 49 **3.69**	2 45 **3.61**	2 42 **3.55**	2 37 **3.45**	2 33 **3.37**	2 28 **3.25**	2 24 **3.18**	2.20 **3.10**	2 16 **3.01**	2 13 **2.96**	2 09 **2.89**	2 07 **2.86**	2 04 **2.80**	2 02 **2.77**	2 01 **2.75**
17	4.45 **8.40**	3 59 **6.11**	3 20 **5.18**	2 96 **4.67**	2.81 **4.34**	2 70 **4.10**	2 62 **3.93**	2 55 **3.79**	2 50 **3.68**	2 45 **3.59**	2 41 **3.52**	2 38 **3.45**	2 33 **3.35**	2 29 **3.27**	2 23 **3.16**	2.19 **3.08**	2 15 **3.00**	2 11 **2.92**	2 08 **2.86**	2 04 **2.79**	2 02 **2.76**	1.99 **2.70**	1 97 **2.67**	1.96 **2.65**
18	4.41 **8.28**	3.55 **6.01**	3.16 **5.09**	2 93 **4.58**	2.77 **4.25**	2.66 **4.01**	2 58 **3.85**	2.51 **3.71**	2 46 **3.60**	2 41 **3.51**	2 37 **3.44**	2 34 **3.37**	2 29 **3.27**	2 25 **3.19**	2 19 **3.07**	2.15 **3.00**	2 11 **2.91**	2 07 **2.83**	2 04 **2.78**	2 00 **2.71**	1.98 **2.68**	1.95 **2.62**	1 93 **2.59**	1 92 **2.57**
19	4 38 **8.18**	3.52 **5.93**	3.13 **5.01**	2 90 **4.50**	2 74 **4.17**	2 63 **3.94**	2.55 **3.77**	2 48 **3.63**	2 43 **3.52**	2 38 **3.43**	2.34 **3.36**	2 31 **3.30**	2 26 **3.19**	2 21 **3.12**	2 15 **3.00**	2 11 **2.92**	2 07 **2.84**	2 02 **2.76**	2 00 **2.70**	1 96 **2.63**	1 94 **2.60**	1.91 **2.54**	1 90 **2.51**	1 88 **2.49**
20	4 35 **8.10**	3 49 **5.85**	3.10 **4.94**	2 87 **4.43**	2 71 **4.10**	2 60 **3.87**	2.52 **3.71**	2 45 **3.56**	2.40 **3.45**	2.35 **3.37**	2.31 **3.30**	2 28 **3.23**	2.23 **3.13**	2 18 **3.05**	2 12 **2.94**	2 08 **2.86**	2 04 **2.77**	1 99 **2.69**	1 96 **2.63**	1.92 **2.56**	1.90 **2.53**	1.87 **2.47**	1 85 **2.44**	1.84 **2.42**
21	4 32 **8.02**	3.47 **5.78**	3 07 **4.87**	2 84 **4.37**	2 68 **4.04**	2.57 **3.81**	2 49 **3.65**	2.42 **3.51**	2 37 **3.40**	2 32 **3.31**	2 28 **3.24**	2.25 **3.17**	2 20 **3.07**	2 15 **2.99**	2 09 **2.88**	2.05 **2.80**	2 00 **2.72**	1 96 **2.63**	1 93 **2.58**	1 89 **2.51**	1 87 **2.47**	1.84 **2.42**	1 82 **2.38**	1 81 **2.36**
22	4 30 **7.94**	3.44 **5.72**	3 05 **4.82**	2 82 **4.31**	2 66 **3.99**	2 55 **3.76**	2 47 **3.59**	2.40 **3.45**	2 35 **3.35**	2 30 **3.26**	2 26 **3.18**	2 23 **3.12**	2 18 **3.02**	2 13 **2.94**	2 07 **2.83**	2 03 **2.75**	1.98 **2.67**	1.93 **2.58**	1 91 **2.53**	1 87 **2.46**	1 84 **2.42**	1.81 **2.37**	1.80 **2.33**	1 78 **2.31**
23	4 28 **7.88**	3.42 **5.66**	3 03 **4.76**	2 80 **4.26**	2 64 **3.94**	2 53 **3.71**	2.45 **3.54**	2 38 **3.41**	2 32 **3.30**	2 28 **3.21**	2 24 **3.14**	2 20 **3.07**	2 14 **2.97**	2 10 **2.89**	2 04 **2.78**	2 00 **2.70**	1.96 **2.62**	1 91 **2.53**	1 88 **2.48**	1 84 **2.41**	1 82 **2.37**	1 79 **2.32**	1 77 **2.28**	1 76 **2.26**
24	4 26 **7.82**	3 40 **5.61**	3 01 **4.72**	2 78 **4.22**	2 62 **3.90**	2.51 **3.67**	2 43 **3.50**	2 36 **3.36**	2 30 **3.25**	2 26 **3.17**	2 22 **3.09**	2 18 **3.03**	2 13 **2.93**	2 09 **2.85**	2 02 **2.74**	1 98 **2.66**	1.94 **2.58**	1 89 **2.49**	1 86 **2.44**	1.82 **2.36**	1 80 **2.33**	1.76 **2.27**	1 74 **2.23**	1.73 **2.21**
25	4 24 **7.77**	3.38 **5.57**	2.99 **4.68**	2 76 **4.18**	2 60 **3.86**	2 49 **3.63**	2 41 **3.46**	2 34 **3.32**	2 28 **3.21**	2 24 **3.13**	2 20 **3.05**	2 16 **2.99**	2.11 **2.89**	2 06 **2.81**	2 00 **2.70**	1.96 **2.62**	1 92 **2.54**	1 87 **2.45**	1 84 **2.40**	1 80 **2.32**	1 77 **2.29**	1 74 **2.23**	1 72 **2.19**	1 71 **2.17**
26	4 22 **7.72**	3.37 **5.53**	2.98 **4.64**	2 74 **4.14**	2 59 **3.82**	2 47 **3.59**	2 39 **3.42**	2 32 **3.29**	2.27 **3.17**	2 22 **3.09**	2 18 **3.02**	2 15 **2.96**	2.10 **2.86**	2 05 **2.77**	1.99 **2.66**	1 95 **2.58**	1 90 **2.50**	1 85 **2.41**	1 82 **2.36**	1 78 **2.28**	1 76 **2.25**	1 72 **2.19**	1.70 **2.15**	1 69 **2.13**

* Table VIII is reproduced from Snedecor: *Statistical Methods*, Iowa State College Press, Ames, Iowa, by permission of the author and publisher.

Table VIII. The 5 (Roman Type) and 1 (Boldface Type) Per Cent Points for the Distribution of F*—Concluded

n_2	n_1 degrees of freedom (for greater mean square)																						
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
50	4 03	3 18	2.79	2.56	2.40	2 29	2 20	2 13	2 07	2 02	1.98	1.95	1 90	1.85	1.78	1.74	1.69	1 63	1 60	1 55	1.52	1 48	1 46	1 44
	7.17	**5.06**	**4.20**	**3.72**	**3.41**	**3.18**	**3.02**	**2.88**	**2.78**	**2.70**	**2.62**	**2.56**	**2.46**	**2.39**	**2.26**	**2.18**	**2.10**	**2.00**	**1.94**	**1.86**	**1.82**	**1.76**	**1.71**	**1.68**
55	4 02	3.17	2.78	2 54	2 38	2.27	2 18	2 11	2 05	2 00	1.97	1.93	1 88	1.83	1.76	1 72	1.67	1.61	1.58	1 52	1.50	1.46	1 43	1 41
	7.12	**5.01**	**4.16**	**3.68**	**3.37**	**3.15**	**2.98**	**2.85**	**2.75**	**2.66**	**2.59**	**2.53**	**2.43**	**2.35**	**2.23**	**2.15**	**2.06**	**1.96**	**1.90**	**1.82**	**1.78**	**1.71**	**1.66**	**1.64**
60	4 00	3 15	2 76	2.52	2 37	2 25	2 17	2 10	2 04	1 99	1.95	1 92	1.86	1 81	1 75	1 70	1 65	1 59	1 56	1 50	1 48	1 44	1 41	1 39
	7.08	**4.98**	**4.13**	**3.65**	**3.34**	**3.12**	**2.95**	**2.82**	**2.72**	**2.63**	**2.56**	**2.50**	**2.40**	**2.32**	**2.20**	**2.12**	**2.03**	**1.93**	**1.87**	**1.79**	**1.74**	**1.68**	**1.63**	**1.60**
65	3 99	3 14	2.75	2.51	2 36	2 24	2 15	2 08	2 02	1 98	1 94	1 90	1.85	1 80	1.73	1.68	1 63	1 57	1 54	1 49	1 46	1 42	1.39	1.37
	7.04	**4.95**	**4.10**	**3.62**	**3.31**	**3.09**	**2.93**	**2.79**	**2.70**	**2.61**	**2.54**	**2.47**	**2.37**	**2.30**	**2.18**	**2.09**	**2.00**	**1.90**	**1.84**	**1.76**	**1.71**	**1.64**	**1.60**	**1.56**
70	3 98	3 13	2 74	2 50	2 35	2 23	2 14	2 07	2 01	1.97	1 93	1 89	1 84	1 79	1 72	1 67	1.62	1 56	1 53	1.47	1.45	1.40	1.37	1.35
	7.01	**4.92**	**4.08**	**3.60**	**3.29**	**3.07**	**2.91**	**2.77**	**2.67**	**2.59**	**2.51**	**2.45**	**2.35**	**2.28**	**2.15**	**2.07**	**1.98**	**1.88**	**1.82**	**1.74**	**1.69**	**1.62**	**1.56**	**1.53**
80	3 96	3 11	2 72	2 48	2 33	2 21	2 12	2 05	1 99	1 95	1 91	1 88	1 82	1 77	1 70	1 65	1.60	1 54	1 51	1 45	1 42	1.38	1 35	1 32
	6.96	**4.88**	**4.04**	**3.56**	**3.25**	**3.04**	**2.87**	**2.74**	**2.64**	**2.55**	**2.48**	**2.41**	**2.32**	**2.24**	**2.11**	**2.03**	**1.94**	**1.84**	**1.78**	**1.70**	**1.65**	**1.57**	**1.52**	**1.49**
100	3.94	3 09	2 70	2 46	2 30	2 19	2 10	2 03	1.97	1.92	1.88	1 85	1 79	1 75	1 68	1 63	1 57	1 51	1 48	1.42	1 39	1.34	1.30	1.28
	6.90	**4.82**	**3.98**	**3.51**	**3.20**	**2.99**	**2.82**	**2.69**	**2.59**	**2.51**	**2.43**	**2.36**	**2.26**	**2.19**	**2.06**	**1.98**	**1.89**	**1.79**	**1.73**	**1.64**	**1.59**	**1.51**	**1.46**	**1.43**
125	3 92	3 07	2 68	2 44	2 29	2 17	2 08	2 01	1 95	1.90	1 86	1 83	1 77	1 72	1.65	1 60	1 55	1 49	1 45	1 39	1 36	1.31	1 27	1.25
	6.84	**4.78**	**3.94**	**3.47**	**3.17**	**2.95**	**2.79**	**2.65**	**2.56**	**2.47**	**2.40**	**2.33**	**2.23**	**2.15**	**2.03**	**1.94**	**1.85**	**1.75**	**1.68**	**1.59**	**1.54**	**1.46**	**1.40**	**1.37**
150	3.91	3 06	2.67	2 43	2 27	2 16	2 07	2 00	1.94	1 89	1.85	1 82	1 76	1 71	1 64	1.59	1.54	1.47	1.44	1 37	1 34	1.29	1.25	1 22
	6.81	**4.75**	**3.91**	**3.44**	**3.14**	**2.92**	**2.76**	**2.62**	**2.53**	**2.44**	**2.37**	**2.30**	**2.20**	**2.12**	**2.00**	**1.91**	**1.83**	**1.72**	**1.66**	**1.56**	**1.51**	**1.43**	**1.37**	**1.33**
200	3 89	3.04	2 65	2 41	2 26	2 14	2 05	1 98	1 92	1 87	1 83	1 80	1.74	1 69	1.62	1.57	1.52	1 45	1 42	1 35	1 32	1 26	1.22	1.19
	6.76	**4.71**	**3.88**	**3.41**	**3.11**	**2.90**	**2.73**	**2.60**	**2.50**	**2.41**	**2.34**	**2.28**	**2.17**	**2.09**	**1.97**	**1.88**	**1.79**	**1.69**	**1.62**	**1.53**	**1.48**	**1.39**	**1.33**	**1.28**
400	3 86	3 02	2.62	2.39	2 23	2 12	2 03	1 96	1 90	1 85	1 81	1 78	1.72	1.67	1.60	1 54	1 49	1.42	1 38	1 32	1 28	1.22	1.16	1.13
	6.70	**4.66**	**3.83**	**3.36**	**3.06**	**2.85**	**2.69**	**2.55**	**2.46**	**2.37**	**2.29**	**2.23**	**2.12**	**2.04**	**1.92**	**1.84**	**1.74**	**1.64**	**1.57**	**1.47**	**1.42**	**1.32**	**1.24**	**1.19**
1000	3 85	3 00	2 61	2 38	2 22	2.10	2 02	1.95	1 89	1.84	1.80	1.76	1 70	1.65	1 58	1 53	1.47	1.41	1 36	1 30	1.26	1.19	1.13	1.08
	6.66	**4.62**	**3.80**	**3.34**	**3.04**	**2.82**	**2.66**	**2.53**	**2.43**	**2.34**	**2.26**	**2.20**	**2.09**	**2.01**	**1.89**	**1.81**	**1.71**	**1.61**	**1.54**	**1.44**	**1.38**	**1.28**	**1.19**	**1.11**
∞	3.84	2.99	2.60	2 37	2 21	2 09	2 01	1 94	1 88	1.83	1.79	1.75	1 69	1 64	1 57	1 52	1 46	1.40	1 35	1.28	1 24	1.17	1.11	1.00
	6.64	**4.60**	**3.78**	**3.32**	**3.02**	**2.80**	**2.64**	**2.51**	**2.41**	**2.32**	**2.24**	**2.18**	**2.07**	**1.99**	**1.87**	**1.79**	**1.69**	**1.59**	**1.52**	**1.41**	**1.36**	**1.25**	**1.15**	**1.00**

* Table VIII is reproduced from Snedecor: *Statistical Methods*, Iowa State College Press, Ames, Iowa, by permission of the author and publisher.

Table IX. The 25, 10, 2.5, and 0.5 Per Cent Points for the Distribution of F*

n_2	P	n_1 Degrees of Freedom (for greater mean square)																	
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	.250	5.83	7.50	8 20	8 58	8 82	8 98	9.10	9.19	9.26	9 32	9.41	9.49	9 58	9.63	9.67	9.71	9 76	9 80	9 85
	.100	39 86	49.50	53.59	55.83	57.24	58.20	58 91	59 44	59 86	60.20	60.70	61 22	61 74	62 00	62 26	62.53	62.79	63 06	63.33
	.025	648	800	864	900	922	937	948	957	963	969	977	985	993	997	1,001	1,006	1,010	1,014	1,018
	.005	16,211	20,000	21,615	22,500	23,056	23,437	23,715	23,925	24,091	24,224	24,426	24,630	24,836	24,940	25,044	25,148	25,253	25,359	25,465
2	.250	2.57	3 00	3 15	3 23	3 28	3 31	3 34	3.35	3 37	3.38	3.39	3.41	3.43	3.43	3.44	3 45	3 46	3 47	3.48
	.100	8 53	9.00	9 16	9.24	9.29	9 33	9.35	9.37	9 38	9.39	9.41	9.42	9 44	9 45	9 46	9.47	9 47	9 48	9.49
	.025	38.51	39.00	39 16	39 25	39 30	39 33	39 36	39.37	39.39	39 40	39 42	39 43	39 45	39 46	39.46	39.47	39.48	39.49	39.50
	.005	198	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	200
3	.250	2.02	2 28	2 36	2 39	2 41	2.42	2.43	2.44	2.44	2.44	2 45	2.46	2.46	2.46	2.46	2.47	2.47	2.47	2.47
	.100	5.54	5 46	5 39	5.34	5 31	5.28	5.27	5 25	5 24	5.23	5.22	5.20	5.18	5.18	5 17	5.16	5.15	5.14	5 13
	.025	17.44	16 04	15.44	15 10	14 88	14 74	14.62	14 54	14.47	14 42	14.34	14.25	14 17	14 12	14 08	14.04	13.99	13.95	13.90
	.005	55.55	49.80	47 47	46 20	45.39	44.84	44 43	44.13	43.88	43.69	43.39	43.08	42.78	42.62	42.47	42.31	42.15	41 99	41.83
4	.250	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2 08	2.08	2 08	2.08	2.08	2.08
	.100	4.54	4.32	4 19	4 11	4.05	4 01	3 98	3.95	3.94	3.92	3.90	3.87	3 84	3 83	3.82	3.80	3.79	3.78	3.76
	.025	12.22	10 65	9 98	9 60	9 36	9.20	9.07	8.98	8 90	8 84	8 75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
	.005	31.33	26.28	24.26	23.16	22.46	21.98	21.62	21.35	21.14	20.97	20.70	20.44	20 17	20.03	19.89	19.75	19.61	19.47	19.32
5	.250	1.69	1 85	1.88	1.89	1.89	1.89	1 89	1.89	1.89	1.89	1.89	1.89	1 88	1 88	1.88	1.88	1.87	1 87	1.87
	.100	4.06	3 78	3 62	3.52	3 45	3.40	3 37	3.34	3 32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
	.025	10.01	8.43	7 76	7 39	7 15	6 98	6 85	6 76	6.68	6.62	6 52	6.43	6.33	6 28	6.23	6 18	6.12	6.07	6 02
	.005	22 78	18.31	16.53	15.56	14.94	14 51	14.20	13.96	13.77	13 62	13.38	13.15	12 90	12.78	12.66	12.53	12.40	12.27	12 14
6	.250	1.62	1 76	1.78	1 79	1 79	1.78	1.78	1.78	1.77	1.77	1.77	1.76	1.76	1.75	1.75	1.75	1.74	1 74	1.74
	.100	3.78	3 46	3.29	3.18	3.11	3.05	3.01	2 98	2.96	2.94	2 90	2 87	2 84	2 82	2.80	2.78	2.76	2 74	2.72
	.025	8.81	7.26	6.60	6 23	5.99	5 82	5.70	5.60	5 52	5 46	5.37	5.27	5 17	5.12	5 07	5.01	4.96	4.90	4.85
	.005	18.64	14.54	12 92	12 03	11 46	11 07	10.79	10 57	10 39	10.25	10.03	9.81	9.59	9.47	9.36	9.24	9.12	9.00	8.88
7	.250	1.57	1.70	1 72	1.72	1 71	1 71	1 70	1.70	1.69	1.69	1 68	1 68	1.67	1.67	1.66	1.66	1.65	1.65	1.65
	.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2 72	2.70	2 67	2 63	2 59	2.58	2 56	2.54	2 51	2.49	2.47
	.025	8.07	6 54	5.89	5 52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4 42	4.36	4 31	4 25	4 20	4.14
	.005	16.24	12.40	10.88	10.05	9 52	9 16	8 89	8 68	8.51	8.38	8 18	7.97	7 75	7.64	7.53	7.42	7 31	7.19	7.08
8	.250	1.54	1 66	1.67	1.66	1 66	1.65	1.64	1.64	1.64	1.63	1.62	1.62	1.61	1.60	1.60	1 59	1 59	1 58	1.58
	.100	3.46	3.11	2.92	2.81	2 73	2.67	2 62	2.59	2.56	2.54	2.50	2.46	2 42	2.40	2.38	2.36	2 34	2.32	2.29
	.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4 36	4 30	4 20	4.10	4 00	3 95	3 89	3 84	3 78	3 73	3.67
	.005	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.01	6.81	6.61	6.50	6.40	6.29	6.18	6.06	5.95

* Table IX is reprinted from Maxine Merrington and Catherine M. Thompson: Tables of percentage points of the inverted beta (F) distribution, *Biometrika*, 1943, **33**, 73–78, by permission of the authors and *Biometrika*.

Table IX. The 25, 10, 2.5, and 0.5 Per Cent Points for the Distribution of F*—Continued

n_2	P	n_1 Degrees of Freedom (for greater mean square)																	
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
9	.250	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.54	1.53	1.53
	.100	3.36	3.01	2 81	2 69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
	.025	7.21	5.71	5.08	4 72	4.48	4.32	4.20	4.10	4 03	3.96	3 87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
	.005	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.23	6 03	5.83	5.73	5 62	5.52	5.41	5.30	5.19
10	.250	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1 53	1.52	1.52	1.51	1.51	1 50	1.49	1.48
	.100	3.28	2 92	2 73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
	.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3 26	3.20	3.14	3.08
	.005	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.66	5.47	5.27	5.17	5.07	4.97	4.86	4.75	4.64
11	.250	1.47	1.58	1.58	1.57	1 56	1.55	1.54	1.53	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.47	1.46	1.45
	.100	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2 05	2.03	2.00	1.97
	.025	6.72	5.26	4.63	4 28	4.04	3.88	3 76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
	.005	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	5.42	5.24	5.05	4.86	4.76	4.65	4.55	4.44	4.34	4.23
12	.250	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.45	1.44	1.43	1.42
	.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
	.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
	.005	11.75	8 51	7.23	6 52	6.07	5.76	5.52	5.35	5.20	5.09	4.91	4.72	4.53	4.43	4.33	4.23	4.12	4.01	3.90
13	.250	1.45	1.55	1.55	1.53	1.52	1.51	1 50	1 49	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.42	1.41	1.40
	.100	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2 20	2.16	2.14	2 10	2 05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
	.025	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
	.005	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94	4.82	4.64	4.46	4.27	4.17	4.07	3.97	3.87	3.76	3.65
14	.250	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.41	1.40	1.39	1.38
	.100	3.10	2.73	2 52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
	.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
	.005	11.06	7.92	6 68	6.00	5.56	5.26	5.03	4.86	4.72	4.60	4.43	4.25	4.06	3.96	3.86	3.76	3.66	3.55	3.44
15	.250	1.43	1.52	1 52	1.51	1.49	1.48	1.47	1.46	1.46	1.45	1.44	1.43	1.41	1.41	1.40	1.39	1.38	1.37	1.36
	.100	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
	.025	6.20	4 76	4.15	3 80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2 58	2.52	2.46	2.40
	.005	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.25	4.07	3.88	3.79	3.69	3.58	3.48	3.37	3.26
16	.250	1.42	1.51	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.44	1.43	1.41	1 40	1.39	1.38	1.37	1.36	1.35	1.34
	.100	3 05	2.67	2 46	2 33	2 24	2.18	2.13	2 09	2.06	2.03	1.99	1.94	1 89	1 87	1.84	1.81	1.78	1.75	1.72
	.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2 99	2.89	2.79	2 68	2.63	2.57	2.51	2.45	2.38	2.32
	.005	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4 38	4.27	4.10	3.92	3.73	3.64	3.54	3.44	3 33	3.22	3.11
17	.250	1 42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1 35	1 34	1.33
	.100	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
	.025	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2 50	2.44	2.38	2.32	2.25
	.005	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25	4.14	3.97	3.79	3.61	3.51	3.41	3.31	3.21	3.10	2.98

*Table IX is reprinted from Maxine Merrington and Catherine M. Thompson: Tables of percentage points of the inverted beta (F) distribution, *Biometrika*, 1943, **33**, 73–78, by permission of the authors and *Biometrika*.

Table IX. The 25, 10, 2.5, and 0.5 Per Cent Points for the Distribution of F^*—Continued

n_2	P	n_1 Degrees of Freedom (for greater mean square)																	
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
18	.250	1.41	1.50	1 49	1.48	1.46	1.45	1.44	1.43	1.42	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.32
	.100	3 01	2 62	2.42	2.29	2 20	2.13	2.08	2 04	2.00	1.98	1 93	1 89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
	.025	5.98	4.56	3.95	3 61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2 32	2.26	2.19
	.005	10.22	7.21	6.03	5 37	4.96	4.66	4.44	4.28	4.14	4.03	3.86	3.68	3.50	3 40	3.30	3.20	3 10	2.99	2 87
19	.250	1.41	1.49	1.49	1.47	1.46	1.44	4.43	1.42	1.41	1.41	1.40	1.38	1.37	1 36	1.35	1.34	1.33	1 32	1.30
	.100	2.99	2.61	2.40	2.27	2.18	2.11	2 06	2 02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1 67	1.63
	.025	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2 82	2 72	2.62	2.51	2 45	2.39	2 33	2.27	2.20	2.13
	.005	10.07	7.09	5 92	5.27	4.85	4.56	4.34	4.18	4.04	3.93	3.76	3.59	3.40	3.31	3 21	3 11	3.00	2.89	2.78
20	.250	1.40	1.49	1.48	1.47	1.45	1 44	1.43	1 42	1.41	1.40	1.39	1.37	1.36	1 35	1 34	1.33	1.32	1.31	1.29
	.100	2.97	2.59	2 38	2 25	2 16	2 09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1 71	1.68	1.64	1.61
	.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
	.005	9.94	6.99	5.82	5.17	4.76	4 47	4.26	4.09	3.96	3.85	3.68	3.50	3.32	3.22	3.12	3.02	2.92	2.81	2 69
21	.250	1.40	1.48	1.48	1.46	1.44	1.43	1.42	1.41	1.40	1.39	1 38	1.37	1.35	1.34	1.33	1 32	1.31	1 30	1.28
	.100	2.96	2.57	2.36	2 23	2.14	2 08	2 02	1.98	1.95	1.92	1.88	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
	.025	5.83	4.42	3 82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2 25	2.18	2 11	2.04
	.005	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4 01	3.88	3.77	3.60	3.43	3 24	3 15	3.05	2.95	2.84	2.73	2.61
22	.250	1.40	1.48	1.47	1.45	1 44	1.42	1.41	1 40	1.39	1.39	1.37	1.36	1.34	1.33	1.32	1 31	1.30	1 29	1.28
	.100	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1 97	1 93	1.90	1.86	1.81	1.76	1.73	1.70	1 67	1.64	1.60	1.57
	.025	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2 84	2.76	2.70	2.60	2.50	2 39	2.33	2.27	2.21	2.14	2.08	2.00
	.005	9.73	6.81	5.65	5.02	4 61	4.32	4.11	3.94	3.81	3.70	3.54	3 36	3 18	3.08	2.98	2.88	2.77	2.66	2.55
23	.250	1.39	1.47	1.47	1.45	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.33	1 32	1.31	1.30	1.28	1 27
	.100	2.94	2.55	2.34	2 21	2 11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1 66	1.62	1.59	1.55
	.025	5.75	4.35	3.75	3.41	3 18	3 02	2 90	2.81	2.73	2.67	2.57	2.47	2.36	2 30	2 24	2 18	2 11	2.04	1.97
	.005	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75	3.64	3.47	3.30	3.12	3.02	2 92	2.82	2.71	2.60	2.48
24	.250	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.36	1 35	1.33	1.32	1.31	1.30	1.29	1.28	1 26
	.100	2.93	2.54	2 33	2 19	2 10	2 04	1.98	1 94	1.91	1.88	1 83	1.78	1 73	1.70	1 67	1 64	1.61	1.57	1.53
	.025	5.72	4.32	3.72	3.38	3.15	2.99	2 87	2.78	2 70	2.64	2.54	2.44	2.33	2 27	2 21	2.15	2.08	2.01	1.94
	.005	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3 83	3.69	3 59	3.42	3 25	3.06	2 97	2.87	2.77	2.66	2.55	2.43
25	.250	1.39	1.47	1.46	1.44	1.42	1.41	1.40	1 39	1.38	1.37	1.36	1.34	1.33	1.32	1.31	1 29	1.28	1.27	1.25
	.100	2.92	2 53	2 32	2 18	2.09	2.02	1 97	1 93	1 89	1.87	1 82	1.77	1 72	1.69	1.66	1.63	1.59	1.56	1 52
	.025	5.69	4.29	3.69	3.35	3.13	2.97	2 85	2.75	2.68	2 61	2.51	2 41	2.30	2.24	2.18	2 12	2.05	1.98	1.91
	.005	9.48	6.60	5 46	4.84	4.43	4.15	3.94	3.78	3.64	3.54	3.37	3.20	3.01	2.92	2.82	2 72	2.61	2.50	2.38
26	.250	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.37	1.35	1.34	1 32	1.31	1.30	1 29	1.28	1.26	1.25
	.100	2 91	2.52	2 31	2.17	2 08	2.01	1.96	1.92	1 88	1.86	1.81	1.76	1.71	1 68	1.65	1.61	1.58	1.54	1.50
	.025	5.66	4.27	3.67	3 33	3 10	2 94	2 82	2 73	2.65	2.59	2 49	2.39	2.28	2.22	2.16	2.09	2 03	1.95	1.88
	.005	9 41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60	3.49	3.33	3.15	2.97	2.87	2 77	2 67	2.56	2.45	2 33

* Table IX is reprinted from Maxine Merrington and Catherine M. Thompson: Tables of percentage points of the inverted beta (F) distribution, *Biometrika*, 1943, **33**, 73–78, by permission of the authors and *Biometrika*.

Table IX. The 25, 10, 2.5, and 0.5 Per Cent Points for the Distribution of F*—Concluded

n_2	P	n_1 Degrees of Freedom (for greater mean square)																		
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
27	.250	1.38	1.46	1.45	1 43	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.33	1 32	1.31	1.30	1.28	1.27	1.26	1.24
	.100	2.90	2.51	2.30	2 17	2 07	2.00	1.95	1 91	1 87	1.85	1.80	1 75	1.70	1.67	1 64	1.60	1 57	1.53	1.49
	.025	5.63	4.24	3.65	3 31	3.08	2 92	2 80	2.71	2 63	2.57	2.47	2 36	2.25	2.19	2.13	2 07	2 00	1 93	1 85
	.005	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56	3.45	3.28	3 11	2.93	2.83	2.73	2.63	2 52	2 41	2.29
28	.250	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.31	1.30	1.29	1 28	1.27	1 25	1.24
	.100	2.89	2.50	2.29	2.16	2.06	2.00	1 94	1.90	1 87	1.84	1.79	1 74	1.69	1.66	1.63	1.59	1.56	1 52	1.48
	.025	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2 69	2 61	2.55	2.45	2.34	2 23	2.17	2.11	2.05	1 98	1.91	1.83
	.005	9.28	6.44	5 32	4.70	4.30	4.02	3.81	3 65	3.52	3.41	3.25	3 07	2.89	2.79	2.69	2.59	2.48	2.37	2 25
29	.250	1.38	1.45	1.45	1.43	1.41	1.40	1.38	1.37	1 36	1.35	1 34	1.32	1 31	1.30	1.29	1 27	1.26	1.25	1.23
	.100	2.89	2.50	2.28	2.15	2.06	1 99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1 51	1.47
	.025	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2 32	2 21	2 15	2 09	2.03	1.96	1 89	1.81
	.005	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3 48	3.38	3.21	3.04	2 86	2 76	2 66	2.56	2 45	2.33	2.21
30	.250	1.38	1.45	1.44	1.42	1.41	1.39	1 38	1.37	1 36	1.35	1.34	1 32	1 30	1 29	1 28	1 27	1 26	1 24	1.23
	.100	2 88	2.49	2.28	2 14	2.05	1.98	1.93	1.88	1 85	1.82	1.77	1 72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
	.025	5.57	4.18	3 59	3.25	3.03	2.87	2 75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1 94	1.87	1.79
	.005	9.18	6.35	5.24	4 62	4 23	3.95	3.74	3.58	3.45	3.34	3.18	3.01	2.82	2.73	2.63	2 52	2.42	2.30	2.18
40	.250	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1 34	1.33	1.31	1.30	1 28	1.26	1.25	1.24	1.22	1 21	1.19
	.100	2 84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1 76	1 71	1 66	1.61	1.57	1.54	1 51	1.47	1.42	1.38
	.025	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2 45	2 39	2 29	2 18	2 07	2.01	1.94	1 88	1 80	1 72	1.64
	.005	8 83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	2.95	2.78	2.60	2.50	2.40	2.30	2.18	2 06	1.93
60	.250	1 35	1 42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1 27	1.25	1.24	1.22	1.21	1.19	1.17	1.15
	.100	2.79	2.39	2.18	2 04	1.95	1 87	1.82	1.77	1.74	1.71	1.66	1 60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
	.025	5 29	3.93	3.34	3 01	2.79	2 63	2.51	2.41	2.33	2.27	2 17	2 06	1.94	1.88	1.82	1 74	1.67	1.58	1.48
	.005	8.49	5.80	4.73	4.14	3.76	3 49	3.29	3 13	3.01	2 90	2.74	2 57	2.39	2 29	2.19	2.08	1.96	1.83	1.69
120	.250	1 34	1 40	1.39	1 37	1.35	1.33	1.31	1.30	1.29	1.28	1.26	1.24	1.22	1.21	1.19	1.18	1 16	1.13	1.10
	.100	2 75	2 35	2.13	1.99	1.90	1.82	1.77	1 72	1.68	1.65	1 60	1 54	1 48	1.45	1.41	1 37	1 32	1.26	1.19
	.025	5.15	3.80	3.23	2 89	2 67	2 52	2 39	2.30	2.22	2 16	2 05	1.94	1.82	1.76	1.69	1 61	1.53	1 43	1.31
	.005	8 18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2 71	2.54	2.37	2 19	2 09	1.98	1 87	1.75	1.61	1.43
∞	.250	1 32	1 39	1 37	1.35	1 33	1 31	1 29	1.28	1.27	1.25	1.24	1.22	1.19	1.18	1.16	1.14	1.12	1.08	1.00
	.100	2 71	2 30	2.08	1 94	1 85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1 24	1.17	1.00
	.025	5.02	3 69	3.12	2 79	2.57	2 41	2 29	2.19	2 11	2.05	1.94	1 83	1.71	1.64	1.57	1.48	1.39	1.27	1.00
	005	7.88	5.30	4.28	3 72	3.35	3.09	2.90	2.74	2 62	2.52	2.36	2 19	2.00	1.90	1.79	1.67	1.53	1.36	1.00

* Table IX is reprinted from Maxine Merrington and Catherine M. Thompson: Tables of percentage points of the inverted beta (F) distribution, *Biometrika*, 1943, **33**, 73–78, by permission of the authors and *Biometrika*.

Table Xa. Significant Studentized Ranges for Duncan's New Multiple Range Test with $\alpha = .10$*

df \ k	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	4.130																	
3	3.328	3.330																
4	3.015	3.074	3.081															
5	2.850	2.934	2.964	2.970														
6	2.748	2.846	2.890	2.908	2.911													
7	2.680	2.785	2.838	2.864	2.876	2.878												
8	2.630	2.742	2.800	2.832	2.849	2.857	2.858											
9	2.592	2.708	2.771	2.808	2.829	2.840	2.845	2.847										
10	2.563	2.682	2.748	2.788	2.813	2.827	2.835	2.839	2.839									
11	2.540	2.660	2.730	2.772	2.799	2.817	2.827	2.833	2.835	2.835								
12	2.521	2.643	2.714	2.759	2.789	2.808	2.821	2.828	2.832	2.833	2.833							
13	2.505	2.628	2.701	2.748	2.779	2.800	2.815	2.824	2.829	2.832	2.832	2.832						
14	2.491	2.616	2.690	2.739	2.771	2.794	2.810	2.820	2.827	2.831	2.832	2.833	2.833					
15	2.479	2.605	2.681	2.731	2.765	2.789	2.805	2.817	2.825	2.830	2.833	2.834	2.834	2.834				
16	2.469	2.596	2.673	2.723	2.759	2.784	2.802	2.815	2.824	2.829	2.833	2.835	2.836	2.836	2.836			
17	2.460	2.588	2.665	2.717	2.753	2.780	2.798	2.812	2.822	2.829	2.834	2.836	2.838	2.838	2.838	2.838		
18	2.452	2.580	2.659	2.712	2.749	2.776	2.796	2.810	2.821	2 828	2.834	2.838	2.840	2.840	2.840	2.840	2.840	
19	2.445	2.574	2.653	2.707	2.745	2.773	2.793	2.808	2.820	2.828	2.834	2.839	2.841	2.842	2.843	2.843	2.843	2.843
20	2.439	2.568	2.648	2.702	2.741	2.770	2.791	2.807	2.819	2.828	2.834	2.839	2.843	2.845	2.845	2.845	2.845	2.845
24	2.420	2.550	2.632	2.688	2.729	2.760	2.783	2.801	2.816	2.827	2.835	2.842	2.848	2.851	2.854	2.856	2.857	2.857
30	2.400	2.532	2.615	2.674	2.717	2.750	2.776	2.796	2.813	2.826	2.837	2.846	2.853	2.859	2 863	2 867	2.869	2.871
40	2.381	2.514	2.600	2.660	2.705	2.741	2.769	2.791	2.810	2.825	2.838	2.849	2.858	2.866	2.873	2.878	2.883	2.887
60	2.363	2.497	2.584	2.646	2.694	2.731	2.761	2.786	2 807	2.825	2.839	2.853	2.864	2.874	2 883	2 890	2.897	2.903
120	2.344	2.479	2.568	2.632	2.682	2.722	2.754	2.781	2.804	2.824	2.842	2.857	2.871	2.883	2.893	2.903	2.912	2.920
∞	2.326	2.462	2.552	2.619	2.670	2.712	2.746	2.776	2.801	2.824	2.844	2.861	2.877	2.892	2.905	2.918	2.929	2.939

* The entries in this table were tabulated and made available by H. Leon Harter.

Table Xb. Significant Studentized Ranges for Duncan's New Multiple Range Test with $\alpha = .05$*

df \ k	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	6.085																	
3	4.501	4.516																
4	3.927	4.013	4.033															
5	3.635	3.749	3.797	3.814														
6	3.461	3.587	3.649	3 680	3.694													
7	3.344	3.477	3.548	3.588	3.611	3.622												
8	3.261	3.399	3.475	3.521	3.549	3.566	3.575											
9	3.199	3.339	3.420	3.470	3.502	3.523	3.536	3.544										
10	3.151	3.293	3.376	3.430	3.465	3.489	3.505	3.516	3.522									
11	3.113	3.256	3.342	3.397	3 435	3.462	3.480	3.493	3.501	3.506								
12	3.082	3.225	3.313	3.370	3.410	3.439	3.459	3.474	3.484	3.491	3.496							
13	3.055	3.200	3.289	3.348	3.389	3.419	3.442	3.458	3.470	3.478	3.484	3.488						
14	3.033	3.178	3.268	3.329	3.372	3.403	3.426	3.444	3.457	3.467	3.474	3.479	3.482					
15	3.014	3 160	3.250	3.312	3.356	3.389	3.413	3.432	3.446	3.457	3.465	3.471	3.476	3.478				
16	2.998	3.144	3.235	3.298	3.343	3.376	3.402	3.422	3.437	3.449	3.458	3.465	3.470	3.473	3.477			
17	2.984	3.130	3.222	3.285	3.331	3.366	3.392	3.412	3.429	3.441	3.451	3.459	3.465	3.469	3.473	3.475		
18	2.971	3.118	3.210	3.274	3.321	3.356	3.383	3.405	3.421	3.435	3.445	3.454	3.460	3.465	3.470	3.472	3.474	
19	2.960	3.107	3.199	3.264	3.311	3.347	3.375	3.397	3.415	3.429	3.440	3.449	3.456	3.462	3 467	3.470	3.472	3.473
20	2.950	3.097	3.190	3.255	3.303	3.339	3.368	3.391	3.409	3.424	3.436	3.445	3.453	3.459	3.464	3.467	3.470	3.472
24	2.919	3.066	3.160	3.226	3.276	3.315	3.345	3.370	3.390	3.406	3.420	3.432	3.441	3.449	3.456	3.461	3.465	3.469
30	2.888	3.035	3.131	3 199	3.250	3.290	3.322	3.349	3 371	3.389	3.405	3.418	3.430	3.439	3 447	3.454	3.460	3.466
40	2.858	3.006	3.102	3 171	3.224	3.266	3.300	3.328	3.352	3.373	3.390	3.405	3.418	3.429	3 439	3 448	3.456	3.463
60	2.829	2.976	3.073	3.143	3.198	3.241	3.277	3.307	3.333	3.355	3.374	3.391	3.406	3.419	3.431	3.442	3.451	3.460
120	2.800	2.947	3.045	3.116	3.172	3.217	3.254	3.287	3.314	3 337	3.359	3.377	3.394	3.409	3 423	3.435	3.446	3.457
∞	2 772	2.918	3.017	3.089	3.146	3.193	3.232	3.265	3.294	3.320	3.343	3.363	3.382	3.399	3.414	3.428	3.442	3.454

* The entries in this table were tabulated and made available by H. Leon Harter.

Table Xc. Significant Studentized Ranges for Duncan's New Multiple Range Test with $\alpha = .01$*

df \ k	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	14.04																	
3	8.261	8 321																
4	6.512	6.677	6.740															
5	5.702	5 893	5.989	6.040														
6	5.243	5.439	5 549	5.614	5.655													
7	4.949	5.145	5 260	5 334	5.383	5.416												
8	4.746	4.939	5.057	5.135	5.189	5.227	5.256											
9	4.596	4.787	4.906	4.986	5.043	5.086	5.118	5.142										
10	4.482	4.671	4.790	4.871	4.931	4.975	5.010	5.037	5.058									
11	4.392	4.579	4.697	4.780	4.841	4 887	4.924	4.952	4.975	4.994								
12	4.320	4.504	4.622	4.706	4.767	4.815	4.852	4.883	4.907	4.927	4.944							
13	4.260	4 442	4.560	4.644	4.706	4.755	4.793	4.824	4.850	4.872	4.889	4.904						
14	4.210	4 391	4.508	4.591	4.654	4.704	4.743	4.775	4.802	4.824	4.843	4.859	4.872					
15	4.168	4.347	4.463	4.547	4.610	4.660	4.700	4.733	4.760	4.783	4.803	4.820	4.834	4.846				
16	4 131	4.309	4.425	4.509	4.572	4.622	4.663	4.696	4.724	4.748	4.768	4.786	4.800	4.813	4.825			
17	4.099	4.275	4.391	4.475	4.539	4.589	4.630	4.664	4.693	4.717	4.738	4.756	4.771	4.785	4.797	4.807		
18	4.071	4.246	4.362	4.445	4.509	4.560	4.601	4.635	4.664	4.689	4.711	4.729	4.745	4.759	4.772	4.783	4.792	
19	4.046	4.220	4.335	4.419	4.483	4.534	4.575	4.610	4.639	4.665	4.686	4.705	4.722	4.736	4.749	4.761	4.771	4.780
20	4.024	4.197	4.312	4.395	4.459	4.510	4.552	4.587	4.617	4.642	4.664	4.684	4.701	4.716	4.729	4.741	4.751	4.761
24	3.956	4.126	4.239	4.322	4.386	4.437	4.480	4.516	4.546	4.573	4.596	4.616	4.634	4.651	4.665	4.678	4.690	4.700
30	3.889	4.056	4.168	4.250	4.314	4.366	4.409	4.445	4.477	4.504	4.528	4.550	4.569	4.586	4.601	4.615	4.628	4.640
40	3 825	3.988	4.098	4.180	4.244	4.296	4.339	4.376	4.408	4.436	4.461	4.483	4.503	4.521	4.537	4.553	4.566	4.579
60	3.762	3.922	4.031	4.111	4.174	4.226	4.270	4.307	4.340	4.368	4.394	4.417	4.438	4.456	4.474	4.490	4.504	4.518
120	3.702	3.858	3.965	4.044	4.107	4.158	4.202	4.239	4.272	4.301	4.327	4.351	4.372	4.392	4.410	4.426	4.442	4.456
∞	3.643	3.796	3.900	3.978	4.040	4.091	4.135	4.172	4.205	4.235	4.261	4.285	4.307	4.327	4.345	4.363	4.379	4.394

* The entries in this table were tabulated and made available by H. Leon Harter.

Table Xd. Significant Studentized Ranges for Duncan's New Multiple Range Test with $\alpha = .005$*

df \ k	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	19.93																	
3	10.55	10.63																
4	7.916	8.126	8.210															
5	6.751	6.980	7.100	7.167														
6	6.105	6.334	6.466	6.547	6.600													
7	5.699	5.922	6.057	6.145	6.207	6.250												
8	5.420	5.638	5 773	5.864	5.930	5.978	6.014											
9	5.218	5.430	5.565	5.657	5.725	5.776	5.815	5.846										
10	5.065	5.273	5.405	5.498	5.567	5.620	5.662	5.695	5.722									
11	4.945	5.149	5.280	5.372	5.442	5.496	5.539	5.574	5.603	5.626								
12	4.849	5.048	5.178	5.270	5.341	5.396	5.439	5.475	5 505	5.531	5.552							
13	4.770	4.966	5.094	5.186	5.256	5.312	5.356	5.393	5.424	5.450	5.472	5.492						
14	4.704	4.897	5.023	5.116	5.185	5 241	5.286	5.324	5.355	5.382	5.405	5.425	5.442					
15	4.647	4.838	4.964	5.055	5.125	5.181	5.226	5.264	5.297	5.324	5.348	5.368	5.386	5.402				
16	4.599	4.787	4.912	5.003	5.073	5.129	5.175	5.213	5.245	5.273	5.298	5.319	5.338	5.354	5.368			
17	4.557	4.744	4.867	4.958	5.027	5.084	5.130	5 168	5.201	5.229	5.254	5.275	5.295	5.311	5.327	5.340		
18	4.521	4.705	4.828	4.918	4.987	5.043	5.090	5.129	5.162	5.190	5.215	5.237	5 256	5.274	5.289	5.303	5.316	
19	4.488	4.671	4.793	4.883	4.952	5.008	5.054	5.093	5.127	5.156	5.181	5.203	5.222	5.240	5.256	5.270	5.283	5.295
20	4.460	4.641	4.762	4.851	4.920	4.976	5.022	5.061	5.095	5.124	5.150	5.172	5.193	5.210	5.226	5.241	5.254	5.266
24	4.371	4.547	4.666	4.753	4.822	4.877	4.924	4.963	4.997	5.027	5.053	5.076	5.097	5.116	5.133	5.148	5.162	5.175
30	4.285	4.456	4.572	4.658	4.726	4.781	4.827	4.867	4.901	4.931	4.958	4.981	5.003	5.022	5.040	5.056	5.071	5.085
40	4.202	4.369	4.482	4.566	4.632	4.687	4.733	4.772	4.806	4.837	4.864	4.888	4.910	4.930	4.948	4.965	4.980	4.995
60	4.122	4.284	4.394	4.476	4.541	4.595	4.640	4.679	4.713	4.744	4.771	4.796	4.818	4.838	4.857	4.874	4.890	4.905
120	4 045	4.201	4.308	4.388	4.452	4.505	4.550	4.588	4.622	4.652	4.679	4.704	4.726	4.747	4.766	4.784	4.800	4.815
∞	3.970	4.121	4.225	4.303	4.365	4.417	4.461	4.499	4 532	4.562	4.589	4.614	4.636	4.657	4.676	4.694	4.710	4.726

* The entries in this table were tabulated and made available by H. Leon Harter.

Table Xe. Significant Studentized Ranges for Duncan's New Multiple Range Test with $\alpha = .001$*

df \ k	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	44.69																	
3	18.28	18.45																
4	12.18	12.52	12.67															
5	9.714	10.05	10.24	10.35														
6	8.427	8.743	8.932	9.055	9.139													
7	7.648	7.943	8.127	8.252	8.342	8.409												
8	7.130	7.407	7.584	7.708	7.799	7.869	7.924											
9	6.762	7.024	7.195	7.316	7.407	7.478	7.535	7.582										
10	6.487	6.738	6.902	7.021	7.111	7.182	7.240	7.287	7.327									
11	6.275	6.516	6.676	6.791	6.880	6.950	7.008	7.056	7.097	7.132								
12	6.106	6.340	6.494	6.607	6.695	6.765	6.822	6.870	6.911	6.947	6.978							
13	5.970	6.195	6.346	6.457	6.543	6.612	6.670	6.718	6.759	6.795	6.826	6.854						
14	5.856	6.075	6.223	6.332	6.416	6.485	6.542	6.590	6.631	6.667	6.699	6.727	6.752					
15	5.760	5.974	6.119	6.225	6.309	6.377	6.433	6.481	6.522	6.558	6.590	6.619	6.644	6.666				
16	5.678	5.888	6.030	6.135	6.217	6.284	6.340	6 388	6.429	6.465	6.497	6.525	6.551	6.574	6.595			
17	5.608	5.813	5.953	6.056	6.138	6.204	6.260	6.307	6.348	6.384	6.416	6.444	6.470	6.493	6.514	6.533		
18	5.546	5.748	5.886	5 988	6.068	6.134	6.189	6.236	6.277	6.313	6.345	6.373	6.399	6.422	6.443	6.462	6.480	
19	5.492	5.691	5.826	5.927	6.007	6.072	6.127	6.174	6.214	6.250	6.281	6.310	6.336	6.359	6.380	6.400	6.418	6.434
20	5.444	5.640	5.774	5.873	5.952	6.017	6.071	6.117	6.158	6.193	6.225	6.254	6.279	6.303	6.324	6.344	6.362	6.379
24	5.297	5.484	5.612	5.708	5.784	5.846	5.899	5.945	5.984	6.020	6.051	6.079	6.105	6.129	6.150	6.170	6.188	6.205
30	5.156	5.335	5.457	5.549	5.622	5.682	5.734	5.778	5.817	5.851	5.882	5.910	5.935	5.958	5.980	6.000	6.018	6.036
40	5.022	5.191	5.308	5.396	5.466	5.524	5.574	5.617	5 654	5 688	5.718	5.745	5.770	5.793	5.814	5.834	5.852	5.869
60	4.894	5.055	5.166	5.249	5.317	5.372	5.420	5.461	5.498	5.530	5.559	5.586	5.610	5.632	5.653	5.672	5.690	5.707
120	4.771	4.924	5.029	5.109	5.173	5 226	5.271	5.311	5.346	5.377	5.405	5.431	5.454	5.476	5.496	5.515	5.532	5.549
∞	4.654	4.798	4.898	4.974	5.034	5.085	5.128	5.166	5.199	5.229	5.256	5.280	5.303	5.324	5.343	5.361	5.378	5.394

* The entries in this table were tabulated and made available by H. Leon Harter.

Table XI. Coefficients for Obtaining the Linear and Quadratic Components of the Treatment Sum of Squares When Treatments are Equally Spaced

		Number of Treatments								
$\sum a^2$	Comparison	1	2	3	4	5	6	7	8	9
2	linear	−1	0	1						
6	quadratic	1	−2	1						
20	linear	−3	−1	1	3					
4	quadratic	1	−1	−1	1					
10	linear	−2	−1	0	1	2				
14	quadratic	2	−1	−2	−1	2				
70	linear	−5	−3	−1	1	3	5			
84	quadratic	5	−1	−4	−4	−1	5			
28	linear	−3	−2	−1	0	1	2	3		
84	quadratic	5	0	−3	−4	−3	0	5		
168	linear	−7	−5	−3	−1	1	3	5	7	
168	quadratic	7	1	−3	−5	−5	−3	1	7	
60	linear	−4	−3	−2	−1	0	1	2	3	4
2,772	quadratic	28	7	−8	−17	−20	−17	−8	7	28

Table XIIa. Table of t for One-Sided Comparisons Between k Treatment Means and a Control for a Joint Confidence Coefficient of $P = 95$ Per Cent*

	k, Number Of Treatment Means (Excluding The Control)								
df	1	2	3	4	5	6	7	8	9
5	2.02	2.44	2.68	2.85	2.98	3 08	3.16	3.24	3.30
6	1.94	2.34	2.56	2.71	2.83	2.92	3.00	3.07	3.12
7	1.89	2.27	2.48	2.62	2.73	2.82	2.89	2.95	3.01
8	1.86	2.22	2.42	2.55	2.66	2.74	2.81	2.87	2.92
9	1.83	2.18	2.37	2.50	2.60	2 68	2.75	2.81	2.86
10	1.81	2.15	2.34	2.47	2.56	2.64	2.70	2.76	2.81
11	1.80	2.13	2.31	2.44	2.53	2.60	2.67	2.72	2.77
12	1.78	2.11	2.29	2.41	2.50	2.58	2.64	2.69	2.74
13	1.77	2.09	2.27	2.39	2.48	2.55	2.61	2.66	2.71
14	1.76	2.08	2.25	2.37	2.46	2.53	2.59	2.64	2.69
15	1.75	2.07	2.24	2.36	2.44	2.51	2.57	2.62	2.67
16	1.75	2.06	2.23	2.34	2.43	2.50	2.56	2.61	2.65
17	1.74	2.05	2.22	2.33	2.42	2.49	2.54	2.59	2.64
18	1.73	2.04	2.21	2.32	2.41	2.48	2.53	2.58	2.62
19	1.73	2.03	2.20	2.31	2.40	2.47	2.52	2.57	2.61
20	1 72	2.03	2.19	2.30	2.39	2.46	2.51	2.56	2.60
24	1.71	2.01	2.17	2.28	2.36	2.43	2.48	2.53	2.57
30	1.70	1.99	2.15	2.25	2.33	2.40	2.45	2.50	2.54
40	1.68	1.97	2.13	2.23	2.31	2.37	2.42	2.47	2.51
60	1.67	1.95	2.10	2.21	2.28	2.35	2.39	2.44	2.48
120	1.66	1 93	2.08	2.18	2.26	2.32	2.37	2.41	2 45
inf.	1 64	1.92	2.06	2.16	2.23	2.29	2.34	2.38	2.42

* Table XIIa is reprinted from C. W Dunnett, A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist Ass*, 1955, **50**, 1096–1121, by permission of the author and the editors of the *Journal of the American Statistical Association*.

Table XIIb. Table of t for One-Sided Comparisons Between k Treatment Means and a Control for a Joint Confidence Coefficient of $P = 99$ Per Cent*

	k, Number Of Treatment Means (Excluding The Control)								
df	1	2	3	4	5	6	7	8	9
5	3.37	3.90	4.21	4.43	4.60	4.73	4.85	4.94	5.03
6	3.14	3.61	3.88	4.07	4.21	4.33	4.43	4.51	4.59
7	3 00	3.42	3.66	3.83	3.96	4.07	4.15	4.23	4.30
8	2.90	3.29	3.51	3.67	3.79	3.88	3.96	4.03	4.09
9	2.82	3.19	3.40	3.55	3.66	3.75	3 82	3.89	3.94
10	2.76	3.11	3.31	3.45	3.56	3.64	3.71	3.78	3.83
11	2.72	3.06	3.25	3.38	3.48	3.56	3.63	3.69	3.74
12	2.68	3.01	3.19	3.32	3.42	3.50	3.56	3.62	3.67
13	2.65	2.97	3.15	3.27	3.37	3.44	3.51	3.56	3.61
14	2.62	2.94	3.11	3.23	3.32	3.40	3.46	3.51	3 56
15	2.60	2.91	3.08	3.20	3.29	3.36	3.42	3.47	3.52
16	2.58	2.88	3.05	3.17	3.26	3.33	3.39	3.44	3.48
17	2.57	2.86	3.03	3.14	3.23	3.30	3.36	3.41	3.45
18	2.55	2.84	3.01	3.12	3.21	3.27	3.33	3.38	3.42
19	2.54	2.83	2.99	3.10	3.18	3.25	3.31	3.36	3.40
20	2.53	2.81	2.97	3.08	3.17	3.23	3.29	3.34	3.38
24	2.49	2.77	2.92	3.03	3.11	3.17	3.22	3.27	3.31
30	2.46	2.72	2.87	2.97	3.05	3.11	3.16	3.21	3.24
40	2.42	2.68	2.82	2.92	2.99	3.05	3.10	3.14	3.18
60	2.39	2.64	2.78	2.87	2.94	3.00	3.04	3.08	3.12
120	2.36	2.60	2.73	2.82	2.89	2.94	2.99	3.03	3.06
inf.	2.33	2.56	2.68	2.77	2.84	2.89	2.93	2.97	3.00

* Table XIIb is reprinted from C. W. Dunnett, A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Ass.*, 1955, **50**, 1096–1121, by permission of the author and the editors of the *Journal of the American Statistical Association.*

Table XIIc. Table of t for Two-Sided Comparisons Between k Treatment Means and a Control for a Joint Confidence Coefficient of $P = 95$ Per Cent*

	k, Number Of Treatment Means (Excluding The Control)								
df	1	2	3	4	5	6	7	8	9
5	2.57	3.03	3.39	3.66	3.88	4.06	4.22	4.36	4.49
6	2.45	2.86	3.18	3.41	3.60	3.75	3.88	4.00	4.11
7	2.36	2.75	3.04	3.24	3.41	3.54	3.66	3.76	3 86
8	2.31	2.67	2.94	3.13	3.28	3.40	3.51	3.60	3.68
9	2.26	2.61	2.86	3.04	3.18	3.29	3.39	3.48	3.55
10	2.23	2.57	2.81	2.97	3.11	3.21	3.31	3.39	3.46
11	2.20	2.53	2.76	2.92	3.05	3.15	3.24	3.31	3.38
12	2.18	2.50	2.72	2.88	3.00	3.10	3.18	3.25	3.32
13	2.16	2.48	2.69	2.84	2.96	3.06	3.14	3.21	3.27
14	2.14	2.46	2.67	2.81	2.93	3.02	3.10	3.17	3.23
15	2.13	2.44	2.64	2.79	2.90	2.99	3.07	3.13	3.19
16	2.12	2.42	2.63	2.77	2.88	2.96	3.04	3.10	3.16
17	2.11	2.41	2.61	2.75	2.85	2.94	3.01	3.08	3.13
17	2.10	2.40	2.59	2.73	2.84	2.92	2.99	3.05	3.11
19	2.09	2.39	2.58	2.72	2.82	2.90	2.97	3.04	3.09
20	2.09	2.38	2.57	2.70	2.81	2.89	2.96	3.02	3.07
24	2.06	2.35	2.53	2.66	2.76	2.84	2.91	2.96	3.01
30	2.04	2.32	2.50	2.62	2.72	2.79	2.86	2.91	2.96
40	2.02	2.29	2.47	2.58	2.67	2.75	2.81	2.86	2.90
60	2.00	2.27	2.43	2.55	2.63	2.70	2.76	2.81	2.85
120	1.98	2.24	2.40	2.51	2.59	2.66	2.71	2.76	2.80
inf.	1.96	2.21	2.37	2.47	2.55	2.62	2.67	2.71	2.75

* Table XIIc is reprinted from C. W. Dunnett, A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Ass.*, 1955, **50**, 1096–1121, by permission of the author and the editors of the *Journal of the American Statistical Association.*

Table XIId. Table of t for Two-Sided Comparisons Between k Treatment Means and a Control for a Joint Confidence Coefficient of $P = 99$ Per Cent*

	k, Number Of Treatment Means (Excluding The Control)								
df	1	2	3	4	5	6	7	8	9
5	4.03	4.63	5.09	5.44	5.73	5.97	6.18	6.36	6.53
6	3.71	4.22	4.60	4.88	5.11	5.30	5.47	5.61	5.74
7	3.50	3.95	4.28	4.52	4.71	4.87	5.01	5.13	5.24
8	3.36	3.77	4.06	4.27	4.44	4.58	4.70	4.81	4.90
9	3.25	3.63	3.90	4.09	4.24	4.37	4.48	4.57	4.65
10	3.17	3.53	3.78	3.95	4.10	4.21	4.31	4.40	4.47
11	3.11	3.45	3.68	3.85	3.98	4.09	4.18	4.26	4.33
12	3.05	3.39	3.61	3.76	3.89	3.99	4.08	4.15	4.22
13	3.01	3.33	3.54	3.69	3.81	3.91	3.99	4.06	4.13
14	2.98	3.29	3.49	3.64	3.75	3.84	3.92	3.99	4.05
15	2.95	3.25	3.45	3.59	3.70	3.79	3.86	3.93	3.99
16	2.92	3.22	3.41	3.55	3.65	3.74	3.82	3.88	3.93
17	2.90	3.19	3.38	3.51	3.62	3.70	3.77	3.83	3.89
18	2.88	3.17	3.35	3.48	3.58	3.67	3.74	3.80	3.85
19	2.86	3.15	3.33	3.46	3.55	3.64	3.70	3.76	3.81
20	2.85	3.13	3.31	3.43	3.53	3.61	3.67	3.73	3.78
24	2.80	3.07	3.24	3.36	3.45	3.52	3.58	3.64	3.69
30	2.75	3.01	3.17	3.28	3.37	3.44	3.50	3.55	3.59
40	2.70	2.95	3.10	3.21	3.29	3.36	3.41	3.46	3.50
60	2.66	2.90	3.04	3.14	3.22	3.28	3.33	3.38	3.42
120	2.62	2.84	2.98	3.08	3.15	3.21	3.25	3.30	3.33
inf.	2.58	2.79	2.92	3.01	3.08	3.14	3.18	3.22	3.25

* Table XIId is reprinted from C. W. Dunnett, A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Ass.*, 1955, **50**, 1096–1121, by permission of the author and the editors of the *Journal of the American Statistical Association*.

Table XIII. Table of Four-Place Logarithms*

N	0	1	2	3	4	5	6	7	8	9	1 2 3	4 5 6	7 8 9
1.0	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374	4 8 12	17 21 25	29 33 37
1.1	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755	4 8 11	15 19 23	26 30 34
1.2	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106	3 7 10	14 17 21	24 28 31
1 3	[illegible]											10	13 16 19	23 26 29
1 4												9	12 15 18	21 24 27
1 5	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014	3 6 8	11 14 17	20 22 25
1 6	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279	3 5 8	11 13 16	18 21 24
1 7	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529	2 5 7	10 12 15	17 20 22
1 8	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765	2 5 7	9 12 14	16 19 21
1 9	2788	2810	.2833	2856	2878	2900	2923	2945	2967	2989	2 4 7	9 11 13	16 18 20
2 0	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201	2 4 6	8 11 13	15 17 19
2.1	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404	2 4 6	8 10 12	14 16 18
2.2	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598	2 4 6	8 10 12	14 15 17
2 3	3617	3636	3655	.3674	3692	3711	3729	3747	3766	3784	2 4 6	7 9 11	13 15 17
2.4	.3802	3820	3838	3856	3874	3892	3909	3927	3945	3962	2 4 5	7 9 11	12 14 16
2 5	3979	3997	4014	.4031	4048	4065	4082	4099	4116	4133	2 3 5	7 9 10	12 14 15
2 6	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298	2 3 5	7 8 10	11 13 15
2 7	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456	2 3 5	6 8 9	11 13 14
2 8	.4472	4487	4502	4518	4533	4548	4564	4579	4594	4609	2 3 5	6 8 9	11 12 14
2 9	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757	1 3 4	6 7 9	10 12 13
3 0	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900	1 3 4	6 7 9	10 11 13
3 1	4914	.4928	4942	4955	4969	4983	4997	5011	5024	5038	1 3 4	6 7 8	10 11 12
3 2	5051	5065	5079	.5092	5105	5119	5132	5145	5159	5172	1 3 4	5 7 8	9 11 12
3.3	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302	1 3 4	5 6 8	9 10 12
3.4	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428	1 3 4	5 6 8	9 10 11
3 5	.5441	5453	5465	5478	5490	5502	5514	5527	5539	5551	1 2 4	5 6 7	9 10 11
3.6	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670	1 2 4	5 6 7	8 10 11
3 7	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786	1 2 3	5 6 7	8 9 10
3 8	5798	.5809	.5821	.5832	5843	5855	5866	5877	5888	5899	1 2 3	5 6 7	8 9 10
3 9	.5911	5922	5933	5944	5955	5966	5977	5988	5999	6010	1 2 3	4 5 7	8 9 10
4.0	.6021	6031	6042	.6053	6064	6075	6085	6096	6107	6117	1 2 3	4 5 6	8 9 10
4.1	.6128	6138	.6149	.6160	6170	6180	6191	6201	6212	6222	1 2 3	4 5 6	7 8 9
4 2	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325	1 2 3	4 5 6	7 8 9
4 3	6335	6345	6355	6365	6375	6385	.6395	6405	6415	6425	1 2 3	4 5 6	7 8 9
4 4	6435	6444	6454	6464	6474	6484	.6493	6503	6513	6522	1 2 3	4 5 6	7 8 9
4 5	.6532	6542	6551	6561	6571	.6580	.6590	6599	6609	6618	1 2 3	4 5 6	7 8 9
4.6	.6628	6637	6646	6656	6665	6675	.6684	6693	6702	6712	1 2 3	4 5 6	7 7 8
4.7	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803	1 2 3	4 5 5	6 7 8
4 8	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893	1 2 3	4 4 5	6 7 8
4.9	6902	6911	.6920	6928	6937	6946	6955	6964	6972	6981	1 2 3	4 4 5	6 7 8
5.0	6990	6998	.7007	7016	7024	.7033	7042	7050	7059	7067	1 2 3	3 4 5	6 7 8
5.1	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152	1 2 3	3 4 5	6 7 8
5 2	7160	7168	7177	7185	7193	.7202	7210	7218	7226	7235	1 2 2	3 4 5	6 7 7
5 3	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316	1 2 2	3 4 5	6 6 7
5 4	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396	1 2 2	3 4 5	6 6 7

* Table XIII is reprinted from D. E Smith, W. D Reeve, and E. L. Morss *Elementary Mathematical Tables*, Ginn and Company, by permission of the authors and publishers.

To obtain the mantissa for a four-digit number, find in the body of the table the mantissa for the first three digits and then, neglecting the decimal point temporarily, add the number in the proportional-parts table at the right which is on the same line as the mantissa already obtained and in the column corresponding to the fourth digit.

Table XIII. Table of Four-Place Logarithms*—Concluded

N	0	1	2	3	4	5	6	7	8	9	1 2 3	4 5 6	7 8 9
5 5	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474	1 2 2	3 4 5	5 6 7
5 6	7482	7490	7497	7505	7513	.7520	7528	7536	7543	7551	1 2 2	3 4 5	5 6 7
5 7	7559	7566	7574	7582	7589	7597	7604	7612	7619	.7627	1 2 2	3 4 5	5 6 7
5 8	7634	7642	7649	7657	7664	7672	.7679	7686	7694	.7701	1 1 2	3 4 4	5 6 7
5 9	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774	1 1 2	3 4 4	5 6 7
6 0	7782	7789	7796	7803	7810	7818	7825	.7832	7839	7846	1 1 2	3 4 4	5 6 6
6 1	7853	7860	7868	7875	7882	7889	.7896	7903	.7910	.7917	1 1 2	3 4 4	5 6 6
6 2	7924	7931	7938	7945	7952	7959	7966	7973	.7980	7987	1 1 2	3 3 4	5 6 6
6 3	.7993	8000	8007	8014	8021	8028	8035	.8041	.8048	8055	1 1 2	3 3 4	5 5 6
6 4	8062	8069	.8075	.8082	8089	8096	.8102	.8109	.8116	.8122	1 1 2	3 3 4	5 5 6
6 5	.8129	8136	8142	8149	8156	.8162	8169	8176	.8182	.8189	1 1 2	3 3 4	5 5 6
6 6	8195	8202	.8209	8215	8222	8228	.8235	8241	.8248	.8254	1 1 2	3 3 4	5 5 6
6 7	8261	8267	.8274	.8280	8287	.8293	8299	.8306	.8312	8319	1 1 2	3 3 4	5 5 6
6 8	8325	8331	.8338	.8344	8351	8357	8363	8370	8376	8382	1 1 2	3 3 4	4 5 6
6 9	8388	8395	8401	.8407	8414	8420	8426	8432	.8439	.8445	1 1 2	2 3 4	4 5 6
7 0	8451	8457	.8463	8470	8476	8482	8488	8494	8500	8506	1 1 2	2 3 4	4 5 6
7 1	8513	8519	8525	.8531	8537	8543	8549	8555	.8561	.8567	1 1 2	2 3 4	4 5 5
7 2	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627	1 1 2	2 3 4	4 5 5
7.3	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686	1 1 2	2 3 4	4 5 5
7.4	8692	8698	8704	.8710	8716	8722	.8727	.8733	.8739	8745	1 1 2	2 3 4	4 5 5
7.5	8751	8756	8762	8768	8774	.8779	8785	.8791	8797	8802	1 1 2	2 3 3	4 5 5
7.6	8808	8814	.8820	8825	8831	.8837	8842	8848	.8854	8859	1 1 2	2 3 3	4 5 5
7 7	8865	8871	.8876	8882	8887	.8893	.8899	8904	8910	8915	1 1 2	2 3 3	4 4 5
7.8	8921	8927	.8932	8938	8943	8949	8954	8960	8965	.8971	1 1 2	2 3 3	4 4 5
7 9	8976	8982	.8987	.8993	8998	9004	.9009	9015	9020	9025	1 1 2	2 3 3	4 4 5
8 0	9031	9036	9042	9047	9053	9058	9063	9069	9074	.9079	1 1 2	2 3 3	4 4 5
8 1	9085	9090	9096	9101	9106	9112	9117	9122	9128	.9133	1 1 2	2 3 3	4 4 5
8 2	9138	9143	9149	9154	.9159	9165	9170	9175	9180	.9186	1 1 2	2 3 3	4 4 5
8 3	9191	9196	9201	.9206	9212	9217	9222	9227	9232	9238	1 1 2	2 3 3	4 4 5
8 4	9243	9248	.9253	9258	9263	9269	.9274	9279	.9284	9289	1 1 2	2 3 3	4 4 5
8 5	9294	9299	9304	.9309	9315	9320	.9325	.9330	9335	.9340	1 1 2	2 3 3	4 4 5
8.6	9345	9350	.9355	9360	9365	9370	9375	9380	.9385	.9390	1 1 2	2 3 3	4 4 5
8 7	9395	9400	.9405	9410	9415	.9420	.9425	9430	.9435	.9440	0 1 1	2 2 3	3 4 4
8 8	9445	9450	9455	9460	9465	9469	9474	9479	.9484	.9489	0 1 1	2 2 3	3 4 4
8 9	9494	9499	9504	.9509	9513	.9518	9523	.9528	.9533	.9538	0 1 1	2 2 3	3 4 4
9 0	9542	9547	9552	9557	9562	9566	9571	9576	.9581	.9586	0 1 1	2 2 3	3 4 4
9 1	9590	9595	9600	9605	9609	9614	9619	9624	.9628	9633	0 1 1	2 2 3	3 4 4
9 2	9638	9643	.9647	.9652	9657	9661	9666	9671	.9675	.9680	0 1 1	2 2 3	3 4 4
9 3	9685	.9689	9694	9699	9703	9708	9713	9717	.9722	9727	0 1 1	2 2 3	3 4 4
9 4	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773	0 1 1	2 2 3	3 4 4
9 5	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818	0 1 1	2 2 3	3 4 4
9.6	9823	9827	9832	9836	9841	9845	9850	.9854	9859	9863	0 1 1	2 2 3	3 4 4
9 7	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908	0 1 1	2 2 3	3 4 4
9 8	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952	0 1 1	2 2 3	3 4 4
9 9	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996	0 1 1	2 2 3	3 3 4

* Table XIII is reprinted from D. E. Smith, W. D. Reeve, and E. L. Morss' *Elementary Mathematical Tables*, Ginn and Company, by permission of the authors and publishers.

To obtain the mantissa for a four-digit number, find in the body of the table the mantissa for the first three digits and then, neglecting the decimal point temporarily, add the number in the proportional-parts table at the right which is on the same line as the mantissa already obtained and in the column corresponding to the fourth digit.

ANSWERS TO PROBLEMS

CHAPTER 2

1. (a) $P = .04$
   (b) $P = .02$
2. (a) $P = 1/70 = .01$
   (b) $P = 16/70 = .23$
   (c) $P = 36/70 = .51$
   (d) 6 Right, 0 Wrong: $P = 1/924 = .0011$
   5 Right, 1 Wrong: $P = 36/924 = .0390$
   4 Right, 2 Wrong: $P = 225/924 = 2435$
   3 Right, 3 Wrong: $P = 400/924 = .4329$
   2 Right, 4 Wrong: $P = 225/924 = .2435$
   1 Right, 5 Wrong: $P = 36/924 = .0390$
   0 Right, 6 Wrong: $P = 1/924 = .0011$
3. (a) $P = 1/4 = .25$
   (b) $P = 12/256 = .05$
   (c) $P = 27/64 = .42$
   (d) 2
   (e) 20
4. 105
7. $P = .03$

CHAPTER 3

2. (a) $P = 1/2$
   (b) 70
   (c) 1/70
   (d) 1/6,720
   (e) 120/6,720
4. (a) .0148
   (b) 3.33
10. (a) 12.0
    (b) 3.0
3. (a) 1/70
   (b) 16/70
   (c) 36/70
   (d) 16/70
   (e) 1/70
5. (a) 0101
   (b) 2.27

CHAPTER 4

1. $z = 1.92$, $P = .0274$
2. $z = 1.79$, $P = .0367$
3. $z = 2.01$, $P = .0222$
4. $z = 1.74$, $P = .0409$
5. $z = 1.75$, $P = .0401$
6. $z = 1.68$, $P = .0465$
7. $z = 1.56$, $P = .0594$
8. $z = 1.88$, $P = (2)(.0301) = .0602$
9. $z = .92$, $P = (2)(.1788) = .3576$
10. $z = 1.54$, $P = (2)(.0618) = .1236$

CHAPTER 5

1. $\chi^2 = 13.16$, $df = 1$, $P < .01$
2. $P = .0002$
3. $\chi^2 = 1.14$, $df = 1$, $P > 20$
4. $\chi^2 = 44\,72$, $df = 2$, $P < .01$
5. $\chi^2 = 54.58$, $df = 9$, $P < .01$
6. $\chi^2 = 23\,82$, $df = 11$, $P < .02$
7. $\chi^2 = 46\,65$, $df = 1$, $P < .01$
8. $\chi^2 = 9\,27$, $df = 1$, $P < .01$

CHAPTER 6

2. .680 and .902
3. $z = .38$, $P = (2)(.3520) = .7040$
4. $z = .595$; $\chi^2 = .354$
5. $\chi^2 = 15.02$, $df = 4$, $P < .01$

CHAPTER 7

1. 20.11 and 24 69
2. $t = 2.80$, $df = 338$, $P < .01$
3. $t = 4.38$, $df = 38$, $P < .01$
4. $t = 3\,01$, $df = 40$, $P < .01$
5. $t = 2.46$, $df = 48$, $P < .05$
6. $t = 1.11$, $df = 24$, $P > .20$
7. $t = 1.74$, $df = 38$, $P > .05$
8. (a) Approximately 28
   (b) Approximately 13

CHAPTER 8

1. (a) $F = 3.16$, $df = 19$ and 9, $P > .05$
   (b) $t = 1.60$, $df = 28$, $P > .10$
2. (a) $F = 4.14$, $df = 19$ and 19, $P < .01$
   (b) $t = 2\,30$ With $\alpha = .05$, the tabled value of t for 19 df is 2.093
3. (a) $F = 3\,32$, $df = 24$ and 24, $P < .01$
   (b) $t = 2\,29$ With $\alpha = 05$, the tabled value of t for 24 df is 2.064
4. (a) $F = 3.27$, $df = 45$ and 46, $P < .01$
   (b) $t = 2\,72$ The approximate value required for significance at the 5 per cent level is 2 01
5. (a) $F = 8\,19$, $df = 107$ and 77, $P < .01$
   (b) $t = 1.59$ The approximate value required for significance at the 5 per cent level is 1.98
6. (a) $F = 2.31$, $df = 69$ and 69, $P < .01$
   $t = 7\,70$ With $\alpha = .05$, the tabled value of t for 69 df is approximately 1.99
7. (a) $F = 1.52$, $df = 46$ and 44, $P > .05$
   $t = 7.03$, $df = 90$, $P < .01$
8. (a) The tabled value for 19 df or $t = 2.093$
   (b) The tabled value for 38 df or $t = 2.025$

CHAPTER 9

1. The mean square for treatments (31.67) is smaller than the mean square for error (82.67) and obviously cannot be significantly larger. There is no need to calculate F.
2. $F = 9.06$, $df = 1$ and 40, $P < 005$
3. The mean square between samples (4.95) is smaller than the mean square within samples (8.60) and obviously cannot be significantly larger. There is no need to calculate F.

4. The mean square for treatments (53.79) is smaller than the mean square for error (93.59) and obviously cannot be significantly larger. There is no need to calculate F
5. $F = 6\ 52$, $df = 4$ and 45, $P < .005$
6. $F = 11\ 09$, $df = 3$ and 60, $P < .005$
7. (a)

Operator	A	B	C	D
$\bar{X}$	7 48	3.11	6 85	6.27
s	3 11	1.24	3.02	2.33

   (b) $\chi^2 = 14.93$, $df = 3$, $P < 01$

   (c)

Operator	A	B	C	D
$\bar{X}$	.901	.585	.865	.839
s	.160	.181	.169	.151

   (d) $F = 13\ 8$, $df = 3$ and 60, $P < 005$
8. (a)

Treatment group	1	2	3
$\bar{X}$	15 30	2.40	4.70
s^2	26.23	4.27	9.12

   (b) $F = 35\ 84$, $df = 2$ and 27, $P < .005$

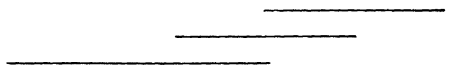
   (c)

Treatment group	1	2	3
$\bar{X}$	3 93	1.57	2.14
s^2	.39	.47	.68

   (d) $F = 29.7$, $df = 2$ and 27, $P < .005$

CHAPTER 10

1. A B C D E F G H



2. For the standard error of the difference, we have

$$s_{\bar{x}_o - \bar{x}_k} = \sqrt{(2)(36)/10} = 2.68$$

With $k = 5$ and d f. $= 54$, for a one-sided test with joint confidence coefficient of $P = 95$ per cent, t is approximately 2.29. For a difference to be significant, we must have

$$\bar{X}_{k\cdot} - \bar{X}_{o\cdot} \geqq (2.68)(2.29) \geqq 6.14$$

With this test, the means for Treatments D and E are significantly greater than the mean of the control group and the means for Treatments A, B, and C are not. For a one-sided test with joint confidence coefficient of $P = 99$ per cent, t is approximately 2.96. Then, for a difference to be significant, we must have

$$\bar{X}_k - \bar{X}_{o\cdot} \geqq (2.68)(2.96) \geqq 7.93$$

and with this test only the mean for Treatment D is significantly greater than the mean for the control group.

5. (a) $F = 4.67$, $df = 2$ and 27, $P < .05$
   (b) $F = 9.0$, $df = 1$ and 27, $P < .01$

CHAPTER 11

1. $F = 22\ 73$, $df = 2$ and 18, $P < 005$
2. $F = 16.26$, $df = 4$ and 20, $P < .005$
3. (a) $F = 5\ 10$, $df = 1$ and 9, $P > .05$
   (b) $t = .6/.266 = 2.26$, $t^2 = 5.1$
4. $F = 2.20$, $df = 2$ and 58, $P > .10$

CHAPTER 12

1. Source of Variation	*F*	*df*	*P*
Type	337.47	1 and 392	<.005
Background	6.17	1 and 392	<.025
Time	988.59	1 and 392	<.005
Type × background	—	—	—
Type × time	155.29	1 and 392	<.005
Background × time	—	—	—
Type × background × time	1.63	1 and 392	>.10

2. Source of Variation	*F*	*df*	*P*
A	3.77	1 and 28	>.05
B	—	—	—
A × *B*	1.60	1 and 28	>.25

3. Source of Variation	*F*	*df*	*P*
A	—	—	—
B	7.39	1 and 56	<.01
C	—	—	—
A × *B*	—	—	—
A × *C*	—	—	—
B × *C*	2.41	1 and 56	>.10
A × *B* × *C*	—	—	—

CHAPTER 13

2. Source of Variation	*F*	*df*	*P*
A	1.34	3 and 48	>.25
B	3.49	2 and 48	<.05
A × *B*	—	—	—

3. Source of Variation	*F*	*df*	*P*
S(ex) of subjects	15.80	1 and 11	<.005
I(nstructions)	—	—	—
B(arrier)	7.22	1 and 11	<.025
E(xperimenter)	6.06	1 and 11	<.05

4. Source of Variation	*df*	Mean Square
A	1	200.00
B	2	405.56
C	2	238.89
A × *B*	2	416.66
A × *C*	2	50.00
B × *C*	4	1,755.56
A × *B* × *C*	4	266.67

5. Source of Variation	*F*	*df*	*P*
A	35.37	1 and 54	<.005
B	290.21	2 and 54	<.005
A × *B*	3.03	2 and 54	>.05

CHAPTER 14

1. (a) Trials: $F = 98.4$, $df = 3$ and 12, $P < .005$
   (b) Linear: $F = 294.5$, $df = 1$ and 12, $P < .005$

2. (a)

Source of Variation	F	df	P
Treatments	—	—	—
Days	81.2	2 and 24	<.005
Treatments × days	2.1	4 and 24	>.10
(b) Linear	157.9	1 and 24	<.005
Quadratic	4.5	1 and 24	<.05
(c) Linear	4.1	2 and 24	<.05
Quadratic	—	—	—

3. (a)

Source of Variation	F	df	P
A	33.45	1 and 8	<.005
B	88.59	1 and 8	<.005
$A \times B$	—	—	—
C	54.37	3 and 24	<.005
$A \times C$	—	—	—
$B \times C$	—	—	—
$A \times B \times C$	1.04	3 and 24	>.250

(b) Linear: $F = 162\,80$, $df = 1$ and 24, $P < .005$

(c) The linear and quadratic components are not significant.

4. (a)

Source of Variation	F	df	P
Instructions	4.67	2 and 18	<.025
Trials	68.83	4 and 72	<.005
Trials × instructions	1.23	8 and 72	>.250

CHAPTER 15

1. (a) $F = 8.87$, $df = 4$ and 12, $P < .005$
   (b) $F = 11.92$, $df = 4$ and 12, $P < .005$
2. Square 1: $F = 18.11$, $df = 3$ and 6, $P < .005$
   Square 2: $F = 2.77$, $df = 3$ and 6, $P > .250$
   Square 3: $F = 26.28$, $df = 3$ and 6, $P < .005$
   Square 4: $F = 2.46$, $df = 3$ and 6, $P > .100$
   For the test of homogeneity of the error mean squares, we have χ^2 (corrected) = 8.97, $df = 3$, and $P < .05$.
3.

Source of Variation	F	df	P
Orders	—	4 and 20	—
Screen size	9.22	4 and 92	<.005
Trials	1.43	4 and 92	>.100

CHAPTER 16

1. (c) F is less than 1.0
   (d) $F = 19.69$, $df = 2$ and 12, $P < .005$
   (e) F is less than 1.0
   (f) $F = 62.75$, $df = 2$ and 11, $P < .005$
   (g) $F = 61.92$, $df = 2$ and 12, $P < .005$
2. (a) F is less than 1.0
   (b) $F = 13\,72$, $df = 3$ and 20, $P < .005$
   (c) $F = 15.34$, $df = 3$ and 19, $P < .005$

AUTHOR INDEX

Anderson, R. L., 302
Archer, E. J., 277

Bancroft, T. A., 302
Bartlett, M. S., 125, 128, 130
Bliss, C. I., 131, 259, 260
Box, G. E. P., 128, 132
Brown, C. W., 1
Bush, R. R., 51, 130

Child, I. L., 220, 221
Clark, M., 115
Cochran, W. G., 57, 63, 107, 162, 212, 228, 272, 273, 276, 303
Cook, S W., 1
Cox, D. R., 212, 296, 297
Cox, G. M., 162, 212, 228, 272, 273, 276
Crespi, L. P., 131
Crump, S. L , 302
Curtiss, J. H., 128

Dallenbach, K. M., 199
DeLury, D. B., 278
Deutsch, M., 1
Duncan, D. B., 137, 138, 140, 198, 257, 294
Dunnett, C. W., 152
Dunnette, M. D., 299

Edwards, A. L., 11, 210

Federer, W. T., 136, 162, 272, 273, 299
Feldt, L. S., 297
Festinger, L., 1
Finney, D. J., 55
Fisher, R. A., 19, 80, 81, 84, 105, 117, 118, 131, 150, 258, 277
Fowler, R. G., 76

Franklin, M., 62
Freeman, M. F., 128
French, E. G., 115

Ghiselli, E. E., 1
Glanville, A. D., 199
Goodman, L. A., 69
Graham, F. K., 115
Grandage, A., 150
Grant, D. A., 233, 245, 277
Guilford, J. P., 131

Haggard, E. A., 130
Harter, H. L., 138
Hartman, G., 75
Hellman, M., 74
Herrera, L., 55
Hoggatt, A. C., 299
Horst, P , 210, 299
Hotelling, H , 85

Jahoda, M., 1

Katz, D., 1
Kempthorne, O., 162, 165, 212, 228, 264, 273, 311
Kendall, B S., 115
Kendall, M. G., 74
Kendler, H. H., 114
Kramer, C. Y , 137, 294
Kreezer, C. L., 199
Kruskal, W. H., 69
Kuenne, M. R., 74

Latscha, R., 55
Lev, J., 307
Lewis, H. B., 62
Lindzey, G., 1

Maier, N. R. F., 69
Mainland, D., 55
McNemar, Q , 57
Merrington, M., 105
Merritt, C. B., 76
Metakides, T., 150
Moore, K., 115
Moore, P. G., 168
Morgan, C. T., 11, 130
Morgan, J. J. B., 101, 132
Mosteller, F., 51, 130
Mowrer, O H., 300
Mueller, C. G., 128

Patel, A. S., 233
Pearce, S. C., 276

Rose, C. L., 259, 260
Rosenzweig, S., 65
Ryan, T. A., 136

Satterthwaite, F. E., 303
Scheffé, H., 154, 156, 191, 206
Schroeder, E. M., 307
Schultz, E. F., Jr., 302, 303, 310
Siegel, S., 11
Sleight, R. B., 129, 131, 277
Smith, B B., 74
Snedecor, G. W., 105, 131, 162, 175, 228, 299
Stevens, S S., 11
Sutcliffe, M. I., 55

Underwood, B. J., 1

Thomas, F. H., 115
Thompson, C. M., 105
Tukey, J. W., 128, 166, 168

Villars, D. S., 302

Walker, H. M., 307
Wilk, M. G., 264, 311
Wilks, S. S , 299
Williams, E. J., 275
Wishart, J., 117, 150
Woodworth, R. S., 10
Worcester, D. A., 115

Yates, F., 131, 150, 258, 277
Young, D. M., 278

SUBJECT INDEX

Additivity, 109
Analysis of covariance, 281–294
  and Latin square designs, 299
  and nonlinear relationships, 285, 294–295
  product sums in the, 281–283, 287–288
  of a randomized blocks design, 299
  randomized blocks design as an alternative to, 296–298
  of a randomized groups design, 281
  test of significance of adjusted treatment mean square, 292–294
  test of significance of differences between group regression coefficients, 291–292
Analysis of variance, 117–118
  of balanced designs, 276
  of a change-over design, 272–273
  of a cross-over design, 272–273
  of difference measures, 295–296
  and expectations of mean squares, 302–303
  of factorial experiments, 175–183, 201–206
  and fixed effects model, 301
  of independently replicated Latin squares, 259–264
  of a Latin square design, 255–257
  and mixed effects model, 301
  multiple comparisons in, 136–156
  and random effects model, 301
  of a randomized blocks design, 160–162, 169–171
  of a randomized groups design, 118–121
  with replications of the same Latin square, 265–268
  of a single Latin square, 255–257
  of a split-plot design, 309–311
  test of significance in, 121, 123
  transformations of scale in, 128–131

Balanced designs, 275–276
Behavioral variables, 6–7
Binomial expansion, 40
Binomial population, 31–33
  applications in research, 41–42
  finite, 34–37
  infinite, 38
  mean of, 32
  shape of, 48
  standard deviation of, 33
  variance of, 33
Binomial probabilities, approximation of, 46–48
Blocks, variables used in forming, 164–165

Carry-over effects, 159, 274–275
Central limit theorem, 111–112
Change-over design, 272–273
Chi square, 63–64, 66, 68, 70, 73, 83–85, 123–128
  and degrees of freedom, 64, 126
  distribution of, 63
  one sample with c classes, 64–65
  table of, 64
  test for homogeneity of correlation coefficients, 83–85
  test for homogeneity of variance, 123–128
  and test of independence, 68–69
  two or more samples, 65–67, 69–70
Coefficient of correlation, 77
Combinations, 17
Comparative experiments, 10
Comparisons, 140–142
  with a control, 152–154
  multiple, 136, 154–156, 191, 198
  orthogonal, 140–148, 189–191, 196–197, 205–206

Comparisons—*continued*
planned, 147
rule for orthogonal, 143–144
standard errors of, 142
test of significance of, 142–143
and unequal *n*'s, 148
Components of variance, 302
Concomitant measure, 281
Confidence coefficient, 89
Confidence interval, 82
Confidence limits, 156
for the correlation coefficient, 82
for the mean, 89–90
for a mean difference, 93–94
and tests of significance, 90
Continuous variable, 44, 87
Control group, 152
Correction, for chi square, 126–127
for discontinuity, 47, 54, 57, 59
Correlated proportions, standard error of difference between, 58–59
test of significance of, 57–60
Correlation, of variance and mean, 128
Correlation coefficients, 77
average value of, 83–84
confidence limits for, 82
sampling distribution of, 77–78
table of significant values of, 79
test of homogeneity of, 83–85
test of significance of, 78–79, 81–83
test of significance of nonindependent, 85
z' transformation for, 79–81
Cross-over design, 272–273
Cumulative distribution, 128
Curvature, significance of, 152

Degrees of freedom, 64, 66–67, 73, 88, 94, 126, 105–106, 182, 201
and chi square, 64, 66, 73–74, 83, 126
and *F*, 105–106
and interaction sums of squares, 182, 201
and mean squares, 120
and *t*, 88, 94
Dependent variable, 10
Difference measures, analysis of variance of, 295–296
Discontinuity, correction for, 47, 54, 57, 59
Distribution, chi square, 63
cumulative, 128
of *F*, 104–105
frequency, 34
normal, 43–46
Poisson, 128
random sampling, 34, 43
rectangular, 86, 112
of *t*, 88–89
of *z'*, 80
Distribution-free tests, 113

Enumeration data, 4
Error mean square, 119, 138, 142, 145, 162, 177, 211–212, 238, 249–250, 268, 272–273, 277, 303-313
in factorial experiments, 176, 211–212, 303, 305
in Latin square designs, 226, 263–264, 268, 272–273, 311–313
method for determining the appropriate, 303
in a randomized blocks design, 162, 165, 171–172, 215, 306–309
in a randomized groups design, 119, 138, 145, 162, 171–172
in a split-plot design, 238, 310
Error sums of squares, linear and quadratic components of, 249–250
pooling of, 123, 211, 249–250, 263–264, 267, 272, 274
Experimental controls, 19–21, 25
Experimental technique, 264
Experiments, 9–11
comparative, 10
and randomization, 21–22

F, distribution of, 104–105
relation to *t*, 146, 171
test for means, 121, 123
test for variances, 105
Factorial experiments, 175, 201–207
advantages of, 197–198
and analysis of variance models, 302–306
expectations of means squares in, 302–306
and fractional replication, 211–212
with organismic factors, 217–219
and orthogonal comparisons, 189–191, 196–197
in a randomized blocks design, 213–214
single replication of, 211, 305
sums of squares in, 177–179, 202–205

Factors, fixed, 176, 301
  levels of, 175
  organismic variables as, 215–217
  random, 301
Finite population, 13, 35–37, 47
  correction factor for, 40
Fixed effects model, 271, 301
  and factorial experiments, 302–306
  and Latin square designs, 311–313
  and randomized blocks designs, 306–309
Fractional replication, 211–212
  and Latin squares, 270–271
Frequency, standard error of, 37, 39–40
  theoretical, 63
Frequency data, 4
Frequency distribution, 34

Graeco-Latin square, 276–277

Heterogeneity of variance, 108–111, 125–128, 131–132
  conditions making for, 108–111
  and nonadditivity, 109
  and nonnormality, 128
  and nonrandom assignment of subjects, 109
  and organismic variables, 110–111
  sensitivity of F test to, 131–132
  and transformations of scale, 128–131
  and t test, 106–107

Independent variable, 10
Infinite population, 13, 38–40, 48
Interactions, 165, 180–183
  calculation of higher-order, 207–210
  graphic representation of, 186–188
  linear component of, 240–243, 248
  meaning of, 184–188, 192–196
  quadratic component of, 243–244, 248
Inverse sine transformation, 130–131

Latin squares, 255
  balanced, 275–276
  and fractional replication, 270–271
  replication of, 259, 265
  the 2×2, 271–274
Latin square designs, 254–255
  and the analysis of covariance, 299
  analysis of variance of, 255–257
  and analysis of variance models, 311–313
  and carry-over effects, 274–275
  expectations of mean squares in, 311–313
  interactions in, 311–312
  randomization in, 258–259
  sums of squares in, 257
  trend analysis in, 268–269
Levels, of factors, 175
  significance, 18–19
Linear component, 239, 247, 269
  of an error sum of squares, 249–250
  of interactions, 240–243, 248
  significance of, 151, 227, 245
Linearity, deviations from, 151–152
Logarithmic transformation, 130, 168, 203

Main effects, 179, 183–184
Matched groups design, 159
Matching problem, 49–51
Mean square, 120
  between groups, 124–125
  expectations of, 302–303
  within groups, 123
Means, of binomial populations, 32
  confidence limits for, 89–90
  confidence limits for a difference between, 93–94
  difference between two, 90–92
  linear functions of, 143
  and multiple comparisons, 136, 154–156, 162, 191, 195
  and multiple range test, 136–139
  orthogonal comparisons of, 141–143
  sampling distribution of, 86–88
  sampling distribution of difference between two, 91
  standard error of, 86, 88
  standard error of difference between two, 92–93, 107
  test of significance of, 89–90
  test of significance of difference between two, 94
Mixed models, 301, 303–306
  and expectations of mean squares, 303–306
  and Latin square designs, 311–313
  and randomized blocks designs, 306–309
  and split-plot designs, 309–311
Models, in the analysis of variance, 176, 271, 301–313
Multiple comparisons, 136, 154–156, 162, 191, 195
Multiple range test, 136–139

Nonadditivity, 109, 165
in a randomized blocks design, 165–168
sum of squares for, 166–167
test of significance for, 167–168
Nonlinear relationships, in the analysis of covariance, 285, 294–295
Nonparametric tests, 113–114
Normal curve, 45, 55
Normal deviate, 44, 46, 53, 57, 59, 81, 87
Normal distribution, 43–46
Normal probability paper, 128
Normality, and heterogeneity of variance, 128
Null hypothesis, 17
alternatives to, 94–97, 105
testing of, 48–49
Number of observations, 97–100, 132, 152, 175
in a randomized groups design, 118

Observations, discrepant, 168
incidental, 1
quantitative, 4
relevance of, 2
systematic, 1
and variables, 3
One-sided tests, 54, 96–97
Ordered variable, 149
Organismic variables, 7–8, 110–111
as factors, 215–217
response-inferred, 9
Orthogonal comparisons, 140–148, 189–191, 196–197, 205–206, 309
in a randomized blocks design, 309
rules for, 143–144
in split-plot designs, 310
of sums, 144–146
Orthogonal polynomials, 150

Parameters, 33–34
Permutations, 16, 23
Poisson distribution, 128
Polynomials, orthogonal, 150
Population, 13
binomial, 31–33
finite, 13, 34–37, 47
infinite, 13, 48
mean of a binomial, 32
standard deviation of a binomial, 33
variance of a binomial, 33
Power, of a test of significance, 94
Probability, definition of, 17
Product sums, in the analysis of covariance, 281–283, 287–288
between groups, 283
total, 282
within groups, 283
Proportions, 35–36
exact test of significance between two, 55–56
sampling distribution of, 35, 46–47
standard error of, 36–37, 39
standard error of difference between two, 53
test of significance between two, 53
test of significance for correlated, 57–60
Protection levels, 140

Quadratic component, 243–245, 248, 269
of an error sum of squares, 249–250
significance of, 245

Random effects model, 301
and expectations of mean squares, 302–305
and Latin square designs, 311–313
and randomized blocks designs, 306–309
Random sample, 28, 30
Random sampling distribution, 34, 43
Randomization, 21–22, 100–101, 228
and experimental controls, 21–22
importance of, 21–22
in Latin square designs, 258–259
in randomized blocks designs, 160
and tests of significance, 70–71
Randomized blocks designs, 158–159, 254
as alternatives to the analysis of covariance, 296–298
and analysis of variance models, 306–309
efficiency of, 172
error mean square in, 171–172, 215
expectations of mean squares in, 306–309
factorial experiments in, 213–214
forming blocks for, 159
nonadditivity in, 165–168
randomization in, 160
sums of squares in, 162–164
test of significance in, 161–162
and trend analysis, 225, 233–234
with two treatments, 169–171
Randomized groups designs, 118
and the analysis of covariance, 281

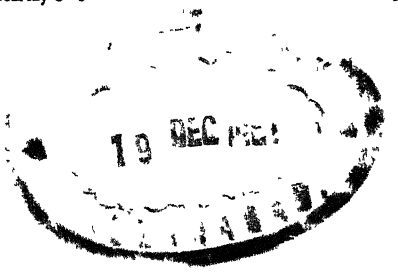
and analysis of variance models, 306–309
between groups sum of squares in, 119
error sum of squares in, 119–120
nature of sums of squares in, 121–123
randomization in, 118
sums of squares in, 119–120
test of significance in, 121
total sum of squares in, 118
within groups sum of squares in, 119–120
Range, shortest significant, 138
studentized, 138
Reciprocal transformation, 131
Rectangular distribution, 86, 112
Regression coefficients, 288–289
estimates of, 290–291
test of significance of differences between, 291–292
Regression line, 288
Replication, 211, 259, 265
Research, 9
Residual sum of squares, 161

Samples, 13, 28, 30,
Sampling distribution, 34, 43
of the correlation coefficient, 77–78, 80
of a difference between two means, 91
of a frequency, 37
of the mean, 86–88
of a proportion, 35
Scale, transformations of, 128–131
Shortest significant range, 138
Significance level, 18–19
Split-plot designs, 228
and analysis of variance models, 309–311
expectations of mean squares in, 309–311
Square root transformation, 128–130, 168
Standard deviation, 88
of a binomial population, 33
Standard error, 36
of a comparison, 153
of difference between two correlated means, 171
of difference between two correlated proportions, 58–59
of difference between two means, 92–93, 107
of difference between two proportions, 53
of difference between two z' values, 82–83
of a frequency, 37, 39–40
of a mean, 86, 88
of an orthogonal comparison, 142
of a proportion, 36–37, 39
of z', 81
Standard normal distribution, 44
Statistics, 33–34
Statistical inference, 14
Stimulus variables, 5–6
Studentized range, 138
Sums of squares, 228–230
between groups, 119
error, 119
in factorial experiments, 177–178, 191–192, 202–205
interaction, 180–182
in Latin square designs, 257
for nonadditivity, 166–167
in a randomized blocks design, 160–164
in a randomized groups design, 121–123
residual, 161
total, 119, 123
within groups, 119–120, 123
Supplementary measures, 281, 298–299

t, distribution of, 88–89
relation to F, 146, 171
t test, 78–79, 89–90, 94
assumptions of, 111–114
and continuity of dependent variable, 113
for the correlation coefficient, 78–79
for the difference between two means, 94
and heterogeneity of variance, 106–107
for the mean, 89–90
and normality of distribution, 111
in a randomized blocks design, 169–171
and skewed distributions, 112
Test of independence, 68–69
Test of significance, 17–18, 53, 145–146, 171
of adjusted treatment mean square in the analysis of covariance, 292–294
in the analysis of variance, 121, 123
and analysis of variance models, 303, 305–306, 308, 310, 312–313
chi square, 64, 66–70
of a comparison, 142–143
and confidence limits, 90
for correlated proportions, 57–60
of correlation coefficient, 81–82
of curvature, 152

Test of Significance—*continued*
of deviations from linearity, 151
of differences between correlation coefficients, 82–83
of differences between means, 94
of differences between proportions, 51–58
of differences between regression coefficients, 291–292
of differences between variances, 105–106, 125–127
distribution-free, 113
of experimental and control group, 154
in factorial experiments, 302–305
in a Latin square design, 256–257, 312–313
of a linear component, 151, 245, 269
multiple range, 136–139
for nonadditivity, 167–168
for nonindependent correlation coefficients, 85
nonparametric, 113–114
one-sided, 54, 96–97
power of, 94
of a proportion, 49
of a quadratic component, 245, 269
and randomization, 70–71
in a randomized blocks design, 161–162
in a randomized groups design, 121
in a split-plot design, 310–311
two-sided, 54, 95–97
of zero correlation, 78–79
Test of technique, 71
Theoretical frequency, 63
Transformations, 79–80, 128–131
logarithmic, 130, 168
inverse sine, 130–131
reciprocal, 131
of scale, 109, 128–131
square root, 128–130, 168
z', 79–80
Treatments, 5–6
Trend analysis, 148–152
in a Latin square design, 268–269
nature of, 224–226
and a randomized blocks design, 225, 233–234
Two-sided test, 54, 95–97, 105
Type I error, 19, 95, 140
Type II error, 17, 25, 43, 94, 97–100, 140

Unit normal distribution, 43–46

Variables, 3–8
behavioral, 6–7
continuous, 4
definition of, 3–4
dependent, 10, 158
in forming blocks, 164–165
independent, 10
ordered, 149
organismic, 7–8, 110–111
qualitative, 5
quantitative, 4–5
response-inferred organismic, 8
stimulus, 5–6
unordered, 5
values of, 3
Variances, 88
of binomial populations, 33
components of, 302
conditions making for heterogeneity of, 108–111
estimate of population, 123–124
heterogeneity of, 106–107, 123–128, 131–132
tests of significance for, 105, 123–128

z', sampling distribution of, 80
standard error of, 81
standard error of difference between two values of, 82–83
transformation, 79–81